JUN 1[illegible] 1921

GENERAL LIBRARY
JUN 21 1921
UNIV. OF MICH.

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS:

HAROLD ORDWAY RUGG, *Chairman.*
Lincoln School of Teachers College.
Teachers College, Columbia University.

JAMES CARLETON BELL,
Brooklyn Training School for Teachers.

FRANK NUGENT FREEMAN,
University of Chicago.

ARTHUR IRVING GATES,
Teachers College, Columbia University.

VIVIAN ALLEN CHARLES HENMON,
University of Wisconsin.

RUDOLF PINTNER,
Ohio State University.

BEARDSLEY RUML,
Carnegie Corporation, New York City.

LEWIS MADISON TERMAN,
Leland Stanford University.

EDWARD LEE THORNDIKE,
Teachers College, Columbia University.

LAURA ZIRBES, *Assistant Editor.*
Lincoln School of Teachers College.

VOL. XII.
No. 5

MAY, 1921

$4.00 a year.
60c. per copy.

CONTENTS

A Survey of the Three First Grades of the Horace Mann School by Means of Psychological Tests and Teachers' Estimates, and a Statistical Evaluation of the Methods Employed. *Clara F. Chassell and Laura M. Chassell* 243

The Scientific Evidence on the Handwriting Movement. *Frank N. Freeman*..... 253

Intelligence and Its Measurement: A Symposium. XIV. By *B. R. Buckingham*..... 271

The True-False Test as a Measure of Achievement in College Courses. *Arthur I. Gates*..... 276

Transmutation of Values on the Thorndike and Ayres Handwriting Scales: A Correction. *T. L. Kelley*... 288

Department for Discussion of Research Problems..... 289

Notes on Articles in Educational Psychology in Current Issues of Other Magazines..... 292

New Publications in Educational Psychology and Related Fields of Education..... 294

Published Monthly Except June to August by
WARWICK and YORK, Inc., Baltimore, Md.

Entered at the Baltimore, (Md.) Postoffice as second-class mail, January 24, 1921.

The Journal of Educational Psychology

is published monthly, except June, July and August, by Warwick & York, Inc., 10 E. Center street, Baltimore, Md. The issues of 1921 make Volume XII.

Manuscripts for publication, books or other materials for review and news items should be addressed to Harold Ordway Rugg, Chairman of the Editorial Board, 59 Edgecliff Terrace, Yonkers, N. Y.

The price of the Journal is $4 a year in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single-issue rate.

Subscribers should notify publishers of changes in address at least four weeks before publication of the issue with which change takes effect.

Educational Psychology Monographs

Studies in the field of educational psychology which are of too great length to be printed in the Journal are issued in book form in the Educational Psychology Monograph Series. These monographs are of varying size and price, and have no definite date for publication. The works listed below have been published, and may be obtained from Warwick & York, Inc., Baltimore, Md.

1. Moto-Sensory Development. George V. N. Dearborn. vi, 215 pages, frontispiece. $1.80.

2. Mental Fatigue. Max Offner; translated by G. M. Whipple. viii, 133 pages. $1.50.

3. Backward and Feeble-Minded Children. Edmund B. Huey. 221 pages, illus. (Out of print.)

4. When Should a Child Begin School. W. H. Winch. 108 pages. $1.40.

5. Spelling Efficiency in Relation to Age, Grade and Sex, and the Question of Transfer. J. E. Wallace Wallin. vi, 91 pages. $1.40.

6. Relative Efficiency of Phonetic Alphabets. Guy Montrose Whipple. 60 pages. 60 cents.

7. Experimental Studies of Mental Defectives. J. E. Wallace Wallin. 156 pages, illus. $1.60.

8. Variations in the Grades of High School Pupils. Clarence Truman Gray. 120 pages. $1.50.

9. How I Kept My Baby Well. Anna G. Noyes. 193 pages and 60 illustrations. $1.80.

10. The Marking System in Theory and Practice. I. E. Finkelstein. 88 pages. $1.40.

11. Inductive Versus Deductive Methods in Teaching. W. H. Winch. 160 pages. $1.80.

12. Children's Perceptions. W. H. Winch. 5-color frontispiece 240 pages. $2.00.

13. The Psychological Methods of Testing Intelligence. W. Stern; translated by G. M. Whipple. 160 pages. $1.60.

14. The Mentality of the Criminal Woman. Jean Weidensall. 350 pages, illus. $2.50.

15. The General Value of Visual Sense Training in Children. Chang Ping Wang. 100 pages. $1.40.

16. The Doctrine of Formal Discipline in the Light of Experimental Investigation. Nellie P. Hewins. 120 pages. $1.50.

17. The Experimental Determination of Mental Discipline in School Studies. Harold O. Rugg. 150 pages. $1.75.

18. The Stanford Revision and Extension of the Binet-Simon Intelligence Scale. Lewis M. Terman and others. 184 pages. $1.80.

19. The Psychology and Pedagogy of Anger. Roy Franklin Richardson. 120 pages. $1.40.

20. The Picture Completion Test. Rudolf Pintner and Margaret M. Anderson. 112 pages, illus. $1.50.

21. Deficiency and Delinquency: An Interpretation of Mental Testing. James Burt Miner. 380 pages. $2.50.

22. The Influence of Age and Experience on Correlations Concerned With Mental Testing. Edward Safford Jones. 98 pages. $1.40.

23. Periodicity With Reference to Mental and Physical Ability. A. G. Peaks. 120 pages. $1.40. (In process.)

24. Graphology and the Psychology of Handwriting. June E. Downey. 180 pages. $1.80.

Warwick & York, Inc., Baltimore, Md.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Vol. XII MAY, 1921 No. 5

A SURVEY OF THE THREE FIRST GRADES OF THE HORACE MANN SCHOOL BY MEANS OF PSYCHOLOGICAL TESTS AND TEACHERS' ESTIMATES, AND A STATISTICAL EVALUATION OF THE METHODS EMPLOYED.

CLARA F. CHASSELL,
Psychologist of the Horace Mann School, Teachers College, Columbia University,

and

LAURA M. CHASSELL,
Instructor in Psychology, Ohio State University.

Part I of this article, published in the February issue of this magazine, contained a statement as to the causes which led to the undertaking of a survey of the three first grades of the Horace Mann School, and presented in full the data obtained in that survey. The measures employed included psychological tests, both group and individual, and rankings of the pupils by the teachers in maturity and ability in reading. In addition, it suggested a plan for utilizing the data thus secured, for the purpose of reclassifying the pupils into relatively homogeneous groups.

Part II continues the report of this survey, recording the correlations obtained between the various measures, evaluating these measures by comparing them with a composite of all the measures utilized, and giving a detailed account of the statistical methods employed in the conversion of these measures into mental ages.[1]

---

[1]For guidance in statistical method and evaluation of measures reported in Part II, the writers are indebted to Prof. T. L. Kelley, formerly of Teachers College, now of Leland Stanford University, and Prof. H. A. Ruger of Teachers College.

PART II. CORRELATIONS, EVALUATION OF MEASURES, AND STATISTICAL PROCEDURE

Presentation of correlations and evaluation of measures.—Table V presents the correlations obtained[2] between the various measures in the survey, including the composite of all. It should be noted that with the exception of the reliability coefficients reported[3] these are based on mental age measures only.

TABLE V

Correlations between Mental Ages Secured by the Various Measures Employed in the Survey

	Stanford Revision	Pressey Primer Scale	Meyer Tests	Combined Group Tests	Teachers' Estimates	Composite
Stanford Revision	(.93)	.45	.53	.64	.72	.89
Pressey Primer Scale	.45	.52[4]	.37			
Meyer Tests	.53	.37	.59[5]			
Combined Group Tests	.64				.42	.87
Teachers' Estimates	.72			.42	(.61)	.85
Stanford Revision and Group Tests						.96
Stanford Revision and Teachers' Estimates						.91
Group Tests and Teachers' Estimates	.77					.98
Composite	.89			.87	.85	

[2]According to the product-moment method of Pearson.

[3]The reliability coefficients actually taken into account in determining the weightings assigned to the various measures entering into the composite, referred to in Part I, were only approximate in the case of the Stanford Revision and the teachers' estimates. The coefficients reported in Table V were variously ascertained. The figure given for the Stanford Revision (.93) is the correlation found between earlier and later tests. (See Terman, L. M., The Intelligence of School Children, p. 142.) The figures for the two group tests (.52 for the Pressey and .59 for the Meyer) were calculated by computing the correlation between the sum of the scores for odd tests and the sum of the scores for even tests. These coefficients are raised to .68 for the Pressey and .74 for the Meyer by the application of 'Brown's' formula. (See Brown, William, The Essentials of Mental Measurement, p. 102, fn. The formula as there stated is $r\,n = \frac{n\,r_1}{1+(n-1)\,r_1}$. In this instance we are not interested in determining the number of applications of the test necessary to give any desired degree of reliability, however, but in determining the reliability of the entire test from the reliability of one-half of the test. Thus, in the case of the Pressey the correlation of .52 between the sum of the scores for odd tests and the sum of the scores for even tests is the reliability coefficient for only one-half of the test. In determining the reliability coefficient of the entire test, $n=2$. Applying the formula, we have $r\,n$, i. e., the reliability coefficient of the entire test, $= \frac{2\,(.52)}{1+(2-1)\,.52} = \frac{2\,(.52)}{1+.52} = .68$. Cf. also the formula given by Spearman, in *British Journal of Psychology*, Vol. III, p. 281. The correlation given for teachers' estimates (.61) is that found to obtain between estimates of school work, made by two or more teachers of the same children in the Horace Mann School, at the time the criterion for the National Research Council Tests was being compiled.

[4]Raised to .68 by the application of 'Brown's' formula.

[5]Raised to .74 by the application of 'Brown's' formula.

An examination of the correlations reported for the Stanford Revision shows that the highest correlation between this measure and any other single measure appears in the case of teachers' estimates, namely, .72 (60 cases being included).[6] In this connection the findings of two other experimenters, while not directly comparable with our results,[7] are sufficiently related to be of interest. Terman[8] reports a correlation of .48 between intelligence quotients and teachers' estimates of intelligence, the rankings being on a scale of 5. Similarly, Dickson[9] found a correlation of .79 between these same measures in the case of 149 first grade children, the teachers having been expressly cautioned to take the children's ages into account in making the ranking.

The correlations with the two group tests[10] are much lower, namely, .45 for the Pressey Primer Scale, and .53 for the Meyer Tests[11] (based on 37 and 45 cases, respectively). A correlation of only .37 was found between the two group tests themselves. A combination of these two tests into a single measure results in a correlation of .64 with the Stanford Revision.

The correlation of .42 between the combined group tests and the teachers' estimates is surprisingly low, since a group examination, to a far greater extent than an individual, would seem to require a response similar to that met by the teachers in the usual classroom situation.

From the practical standpoint, the correlation of greatest interest is probably the one between the Stanford Revision and group tests combined with teachers' estimates. The correlation of .77 between these measures, based on only 30 cases, is sufficiently high to raise the question as to whether an individual psychological examination is really necessary for determining classification and promotion. If the possibility of partially explaining this relatively high correla-

---

[6]It should be borne in mind, however, that the teachers were already familiar with the intelligence quotients of a large number of their children. While it is probable that no direct comparison with these quotients was made, this previous knowledge doubtless had some influence upon the rankings assigned.

The rankings made by the teachers in our survey were of maturity and ability in reading, and in general extended over the entire range of the class.

[8]The Measurement of Intelligence, p. 75.

[9]See Terman, L. M., The Intelligence of School Children, pp. 51-52.

[10]All correlations reported with the combined group tests are based on 30 cases only. This small number of cases in which direct comparison is possible is due to the prolonged absence of an unusually large number of children.

[11]As explained in Part I, the Meyer Tests used in this survey are those constructed by Miss Helen Meyer, described in an unpublished thesis, "Group Tests for Grades I and II," on file in the Columbia University Library.

tion by the fact that the teachers were familiar with the intelligence quotients of a large number of their children be entirely overlooked, it should still be emphasized that a correlation of .77 means a reduction in the ratio of the variability around the regression line to the variability around the average of only .64.[12] Even if this reduction in variability were much greater, in view of the great value of the Stanford Revision for the understanding of individual pupils, we should still hesitate to say that it was not an essential instrument for satisfactory classification and promotion. Certainly, however high this correlation may be found by other investigators, the Stanford Revision or its equivalent will continue to be indispensable at least for determining the placing of the problematical cases of any kind.

Aside from such information as may be afforded by the correlations with the Stanford Revision, already reported, for an evaluation of the various measures used in the survey we are dependent upon a comparison of these measures with the composite of all.[13] As explained in Part I, this composite is the average mental age found by adding the mental ages obtained from the Stanford Revision, the two group tests, and the teachers' estimates, the Terman mental age being doubled, and dividing this total mental age by the number of measures available for each child, the Terman mental age being counted as two measures.

As judged from the correlations with this composite, presented in Table V, any one of the three types of measures which were used in the survey, that is, an individual examination or two group tests or two teachers' estimates, would approximate the same results for classification and promotion as the composite itself. The correlations for these three measures, based on 60, 30, and 60 cases, respectively, are .89, .87, and .85. In interpreting these correlations, as well as the other correlations to be reported with the composite, it must be borne in mind that the measures compared are not independent, but are already contained within the composite itself.

---

[12]Obtained by the formula, $\sqrt{1-r^2}$.

[13]It is impossible to evaluate this composite by comparing it with an altogether independent criterion of the value of the various measures employed for purposes of classification and promotion. Such an independent measure would be available subsequently if the results of instruction in groups classified on the basis of the composite could be compared with results obtained in control groups not so classified. One evidence that the composite employed is a fairly satisfactory criterion by means of which to evaluate the measures used in the survey, however, is the general satisfaction felt by the teachers with the results of the survey as summarized in this composite.

Hence they are far higher than would otherwise be the case.[14] Even so, taking these correlations at their face value, since the ratio of the variability around the regression line to the variability around the average would be reduced approximately only one-half[15] by the use of any one of the three types of measures, taken alone, they can be claimed, if used singly, to possess only a limited value for purposes of classification and promotion as compared with the composite.

The correlations with the composite are naturally very much higher still when any two types of measures are combined for purposes of comparison with it, on account of the fact that the four elements included in the two types of measures (e.g., the Terman mental age taken twice plus the mental ages obtained from the two group tests) are identical with four out of the six elements in the composite. Thus the Stanford Revision combined with the group tests gives a correlation of .96 with the composite, the Stanford Revision combined with the teachers' estimates, of .91, and the group tests combined with the teachers' estimates, of .98,[16] (30 cases being included in each instance). The ratio of the variability around the regression line to the variability around the average for these correlations is, respectively, .28, .41, and .20. In interpreting these results, it should be noted that the presence of common elements in the measures correlated, which served in the first place to increase the correlations with the composite, results now in correspondingly lower regression values. Should the special conditions under which these correlations were obtained remain, if only two types of measures are to be used for purposes of classification and promotion, it would seem from these results to make little difference whether the Stanford Revision combined with group tests or group tests combined with teachers' estimates were selected.

---

[14]The reader who is interested in following up the implications of this statement statistically, may be referred to Yule, G. U., An Introduction to the Theory of Statistics, ch. XI. (See especially Exercises 6 and 7, p. 227.)

[15]Exactly one-half (i. e., 50) if the correlations were all .866. The exact regression values obstined by the formula, $\sqrt{1-r^2}$ for these three correlations are .46, .49 and .53, respectively. It is apparent from this formula, however, that since these regression values involve the correlations with the composite, they are too low for the same reason that the correlations upon which they are based are too high.

[16]This last-mentioned correlation may have been artificially increased by still another factor, since as already stated, the teachers' judgments probably were influenced to a certain extent by previous knowledge of the intelligence quotients of the children. The extent of such influence can not be determined. Whatever it may have been, however, it did not succeed in raising the correlation between the Stanford Revision and the teachers' estimates above .72.

The statistical procedure utilized in the conversion of the measures into mental ages.—Reference has already been made in Part I[17] to the fact that the incorporation of such varied data as were secured in the survey necessitated the reduction of all measures to a common basis. The one selected for this purpose was that of mental age.

Before concluding the report of the survey it is thus necessary to present in detail the statistical procedure involved in converting the various measures used into mental ages. The methods used are given in order for the measures employed.

1. The Stanford Revision.

In the case of the Stanford Revision the only computation necessary was that required on account of the fact that the giving of the individual examinations had extended over a long period. It was thus necessary to determine the mental ages of all the children at some specified time. Since the group tests had been given during that same month, the date selected was January 1. After the chronological ages of the children on that date had been computed, the corresponding mental ages were readily obtained by using the intelligence quotients already determined as multipliers.[18] This process was facilitated by the use of an I. Q. slide rule.[19]

2. The Pressey Primer Scale and the Meyer Tests.

The method used in converting the scores made in both the Pressey Primer Scale and the Meyer Tests into mental ages, consisted of replacing the score made by a given child in one of these tests by the corresponding mental age, as shown in a table constructed with the median scores in the test for the various ages as a basis.

Since only year norms were available for the Pressey tests and no norms whatever for the Meyer, before the conversion into mental ages was possible it was necessary to build up a table of norms by months. In the case of the Pressey the procedure was as follows: First, since the various age norms provided for the Primer Scale actually represent the typical performances of children six months in advance of the ages specified, the norm for any given year being

---

[17]See *Journal of Educational Psychology*, Vol. XII, No. 2 (Feb., 1921), p. 74.

[18]This method is based on the assumption of the constancy of the I. Q., which, although not thoroughly established, seemed to be sufficiently so for use in the present instance.

[19]Published by the Reed College Co-Operative Store, Portland, Oregon.

based upon the scores made by all the children who have passed the birthday indicated and have not yet reached the succeeding birthday, the median score given as the norm for six years of age, for example, was taken to have a value in terms of mental age of six and one-half years. The amount of the interval between each of the median scores for the succeeding ages was then determined, and this amount divided by twelve in order to obtain the increment in terms of score which might be taken to correspond to a month of mental age. Each increment was then successively added to the appropriate year norm. The values thus secured were subsequently entered in a table opposite the equivalent year and month of mental age.

The procedure followed in the case of the Meyer Tests was naturally more complicated. No norms being available, it was necessary to estimate norms by means of a comparison of scores made by children in these tests with scores made by the same children in some other test. For this purpose the records from the Stanford Revision were utilized.

First, the mental ages of the twenty-nine children given the six Meyer tests in the regular manner,[20] all of whom had also been given the Stanford Revision, were arranged according to the respective chronological ages. Then for each successive half year of chronological age the median of the chronological ages and of the mental ages was found. Each median mental age was then divided by the median chronological age corresponding, and the median intelligence quotient for the children of each half year of chronological age found. The scores made by these same children in the Meyer Tests were similarly tabulated according to the chronological ages of the children, and the median score for each half year ascertained. The age medians thus secured were next divided by the intelligence quotients obtained as described above, in order to secure medians which would be typical of children in general, and which thus might serve as norms for the corresponding chronological ages. The values resulting were then employed in the same way in which the Pressey age norms had been utilized, and a table of scores with their

---

[20]Sometime before plans had been made for the survey, two out of the six tests in the Meyer series had been given to the children in one of the rooms. Although, after the remaining four tests had been given to these children, their scores were adapted to make them comparable with those made by the other children, these records were not utilized in estimating the norms.

accompanying equivalents in terms of years and months of mental age built up.

3. The teachers' estimates.

The conversion of the teachers' estimates into mental ages was made by adapting two different methods. Only the general outlines of the methods will be indicated here; the adaptations employed were made in the course of the practical application of the procedures, and would not be of general interest. Both methods involved the transmutation of a given teacher's rankings in each trait, into an equivalent mental age, determined on the basis of the Terman mental ages on January 1st, already calculated, on the principle that for any group of children the distribution of mental age indicated by any measure would be similar to that already found by the application of the Stanford Revision. The first method, which is based on the assumption that the form of distribution concerned is that of the normal probability surface, was employed in the case of the two classes for which all or practically all of the Stanford Revision records were available; the second method, which makes no such assumption, was employed in the case of the third class because the data were relatively incomplete on account of the prolonged absence of a number of the children. For purposes of comparison a third method, used in the survey made in the spring throughout the elementary school, in which the ranks assigned were treated as gross scores, is included.

Method A, in which the distribution was assumed to follow the normal curve. The average and the standard deviation from the average of the Terman mental ages for the class concerned were first computed. Then the number of children assigned each of the teacher's rankings[21] was determined, and the percentage that this number represented of the total number of the children in the group computed. The average distance of each percentage from the central tendency of the total group in terms of multiples of the standard deviation was then ascertained by reference to Table 54 in Thorndike's Mental and Social Measurements.[22] The multiples thus obtained were multiplied by the value of the standard deviation from

[21] In a number of instances more than one child had been assigned the same rank.
[22] See p. 221 ff.

the average of the Terman mental ages already obtained. The resulting quantities were then added algebraically to the average Terman mental age for the class, and the resulting figures replaced in each case by the nearest whole number. This figure, which represented the number of months of mental age desired, was then converted into years and months.

Method B, in which the distribution was treated as rectangular. In the first place, the highest rank was replaced by the highest mental age found for the children concerned, as its equivalent; and, similarly, the lowest rank was replaced by the lowest mental age. The range in months between the highest and the lowest age was then divided by (n—1) the number of ranks, in order to determine the number of months of mental age which should be taken as equivalent to the interval between two successive ranks. The amount thus secured was subtracted from the highest mental age to secure the value of the second highest rank; then subtracted from this value to secure the value of the third highest rank, and so on, until a value had been found for each rank. The quantities resulting were in each case replaced by the nearest whole numbers, and these numbers, which represented the number of months of mental age appropriate to each rank, were then reduced to years and months.

Method C, in which ranks assigned were treated as gross scores. In the first place, since records for identical children in both measures, only, were usable, records for any children who had not been included in the teachers' rankings and examined by the Stanford Revision as well, were eliminated from the calculations. Next, the remaining ranks being treated as gross scores, the average and the standard deviation from the average of these ranks were computed. Similarly, the average and the standard deviation from the average of the Terman mental ages for the same children were found as in the first method described above. Then, the standard deviation of the ranks being considered as equivalent to the standard deviation of the mental ages, the number of months of mental age equivalent to the interval between two succeeding ranks was found. Thus, if the standard deviation of the ranks of a given class was 3, and the standard deviation of the mental ages for the same children was 6 months, the interval between two succeeding ranks would be considered equivalent to two months of mental age. Finally, with the

average mental age as the starting point, the appropriate number of months was added or subtracted for each following or preceding rank; the resulting figures, which represented mental ages, were replaced by the nearest whole numbers, and the latter reduced to years and months.

THE SCIENTIFIC EVIDENCE ON THE HANDWRITING MOVEMENT

FRANK N. FREEMAN,
The University of Chicago.

The Arm Movement Dogma.

The teaching of handwriting is dominated by a very widespread dogma concerning the best way to write and the best way to teach writing. This dogma is the formulation of the opinion that the arm movement, or the so-called muscular movement, is a superior method of writing, and that writing should be taught by emphasizing this arm movement by giving exercises which develop it and fixing the child's attention upon it. Few writing supervisors question the correctness of this dogma. School administrators, in general, accept the dictum of the supervisors. Some teachers, however, have come to question it, as a result of their firsthand experience in trying to teach this movement. Scientific evidence refutes it almost completely. Even casual observation, if it were made without bias, would seriously undermine it.

The strongest refutation of the arm movement dogma is the fact that with all the efforts which are made to teach it, the pupils do not, in the main, acquire it. It is almost a flat failure. An investigation[1] of 273 children, most of whom had been given strenuous drill in arm movement writing, brought to light the proportion of those who acquired the movement with enough thoroughness to be used by them in their ordinary writing. The analysis of the writing movement was made by means of an instrument which gave a tracing of the movement of the arm. The amount of arm movement could be estimated from this tracing according to the degree to which it corresponded to the original writing. The degree of correspondence was estimated by comparison with a scale of five specimens. The first specimen of this scale is designated by zero and represents no arm movement whatever. The specimens from 1 to 5 represent varying degree of arm movement, 5 being the highest. In specimen 3 the

[1] The data referred to in this article are presented in detail in the author's monograph, "The Handwriting Movement," Chicago, 1918, The University of Chicago Press.

arm movement is sufficiently pronounced to cause the hand tracing to resemble somewhat the writing, but not sufficient to enable it to be read.

The results of this study have been tabulated by ages. We may take the data for ages 8 and 14 as representative. These are shown in Table I.

TABLE I.

		Grade of Arm Movement.					
Age.	No.	0	1	2	3	4	5
8	33	.30	.58	.12			
14	38	.03	.42	.26	.11	.11	.08

It will be seen that practically none of the eight-year children used more than the very most rudimentary type of arm movement. Grade I of the scale represents simply a slight upward and downward oscillation of the arm. Even Grade 2 represents a very slight amount. These figures indicate that it is almost out of the question to expect to develop the arm movement in primary grade children.

The showing of the fourteen year old children is not much better. Seventy-one per cent exhibited less arm movement than is represented by Grade 3. Practically none exhibited complete arm movement.

The gist of the matter may be put in more concrete form by a quotation from one of the eighth grade pupils who was a subject in the experiment. This pupil was asked to write under a motion picture camera. In order that she might know what was wanted, she asked, "Do you want my ordinary writing or my arm movement writing?" The pupil did not use the term "arm movement," but used the name of a prominent writing system which emphasizes this type of movement. It is quite evident that this style of writing had not become so habitual with this pupil that it could be used without conscious effort.

The Problem.

This example may serve to illustrate the function which the scientific study of the handwriting movement has to perform. Its purpose may be thought of, first, as the examination of the current assumptions concerning methods of learning and of teaching. Its next task is to determine what the essentials of good writing are

apart from the presuppositions of current practice. The procedure which was followed in the experiment here described was to make an analysis of the behavior of good writers and poor writers, and to endeavor from this comparison to determine what characteristics are essential and what are neglible. The elements of the writing movement which were included in the study comprise position, the gross composition of the movement, and the detailed changes in the speed of the movement.

The Method.

The method consisted of photographing the movement of the hand and arm by a kinetoscopic camera. The camera was mounted directly above the hand of the writer and was speeded up to twenty-five exposures per second. This gave a separate photograph of the hand and of the preceding writing each twenty-fifth of a second. This record was sufficiently detailed for the purposes in hand.

The position of the hand at any point in the writing could be determined by putting the developed film in a projection machine, throwing the image on the screen with the machine still, and making a drawing from it. The nature of a gross movement could be determined by making a succession of these drawings to represent the positions at successive points in the writing. The most detailed examination of the movement was made by plotting the position of certain reference points on the hand at each successive exposure of the camera.

The speed changes in the movement were examined by recording the position of the pen point on the letter strokes at each successive exposure of the camera. From this record of the positions of the pen point at twenty-fifth of a second intervals, it was possible to determine the distance traveled by the pen point in successive units of time. These successive distances were plotted on a speed curve as is illustrated in Figures 3 and 4.

Importance of Position.

The methods of teaching handwriting in common use are similar to the methods of teaching other manual arts in that they assume that certain positions are most favorable to the successful prosecu-

tion of the movement and attempt to teach the pupils to assume these positions habitually. These elements of position comprise a large part of the "good form" of the activity. The accepted features of position in writing have been very little changed for many years. The chief modifications within the last two or three decades have consisted in slight relaxation in the rigidity of the requirements.

Standard position illustrated in Figure 1 by the drawing from the hand of the late Mr. C. P. Zaner, one of the best known writing specialists and teachers. This position, however, deviates slightly from the rigid regulations already alluded to. The chief departure consists in turning the hand slightly over to the right. As a consequence of this the penholder points somewhat farther to the right than is prescribed. Furthermore, the penholder is allowed to rest in the depression betwen the thumb and forefinger, instead of crossing the knuckle joint at the base of the forefinger. The hand rests upon the nails of the third and fourth fingers, however, and the grasp of the pen corresponds quite accurately with the usual directions.

This fairly orthodox type of position may be compared with the hand position of another writer which departs much more widely from the position which is taught. See Figure 2. This is a good writer in the eighth grade. While the wrist is held fairly level, the pen is turned over to the right and the thumb and forefinger are drawn in more than would ordinarily be considered desirable. In spite of this, the writing of this girl is among the best to be found in the eighth grade. Other figures might be shown of individuals whose position is beyond criticism, but whose writing is very poor.

From a small number of such cases one might be tempted to conclude that position is of no importance whatever for good writing. A more comprehensive survey of a larger number of cases shows that position is of some value. The variations which appear among good writers, however, show that more deviation from the rigidity of the orthodox position may be permitted than is common. Our problem is to determine what the essential elements are and to distinguish them from the non-essential.

The more important characteristics of position may be gathered from an examination of Tables II and III. Table II is the key to the understanding of Table III. Hand position was first examined with reference to the extent to which the hand is turned down so that the wrist is level. This is called "degree of pronation." The method

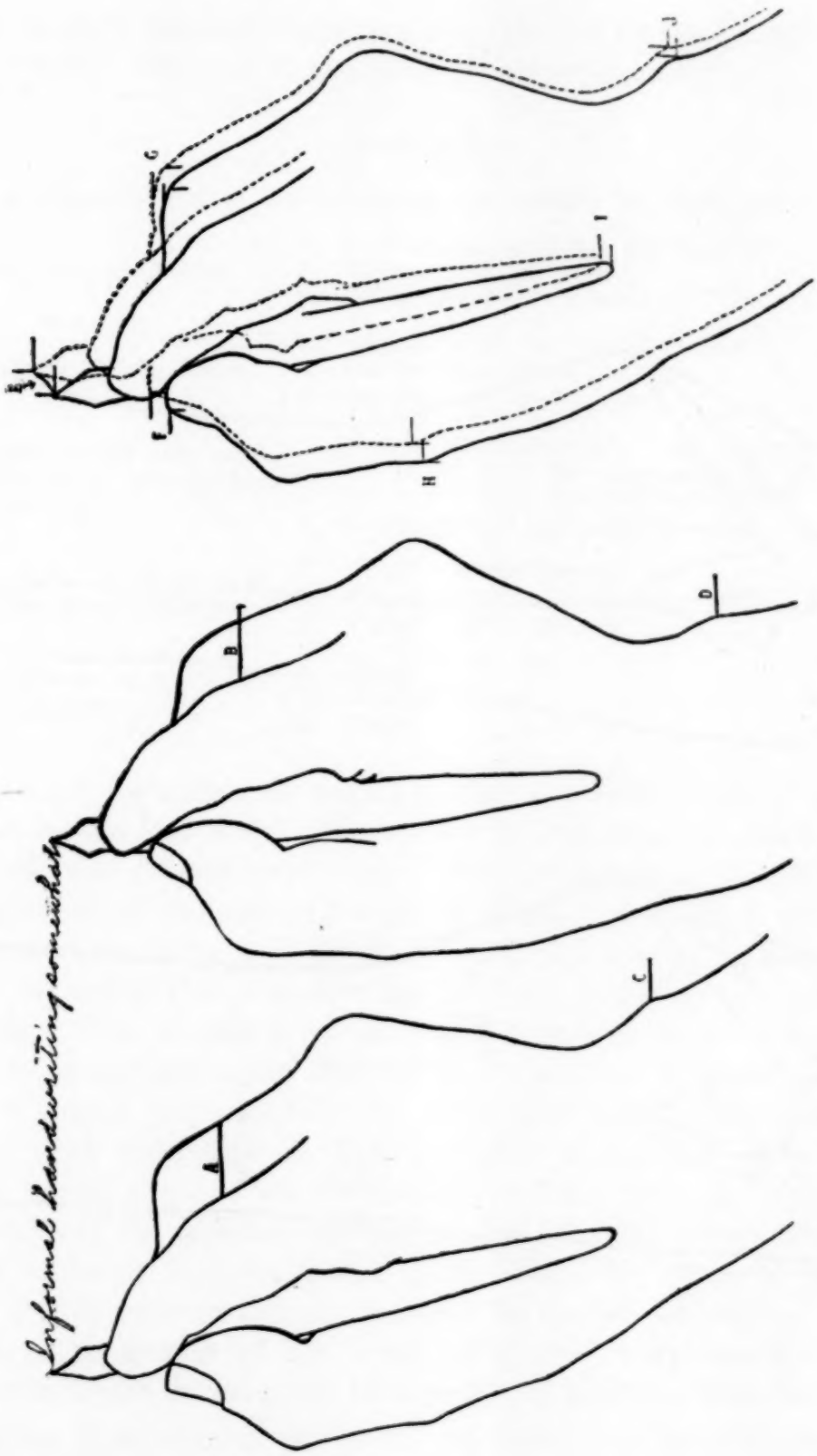

FIG. 1.—Illustration of the drawings showing hand position and gross movement.

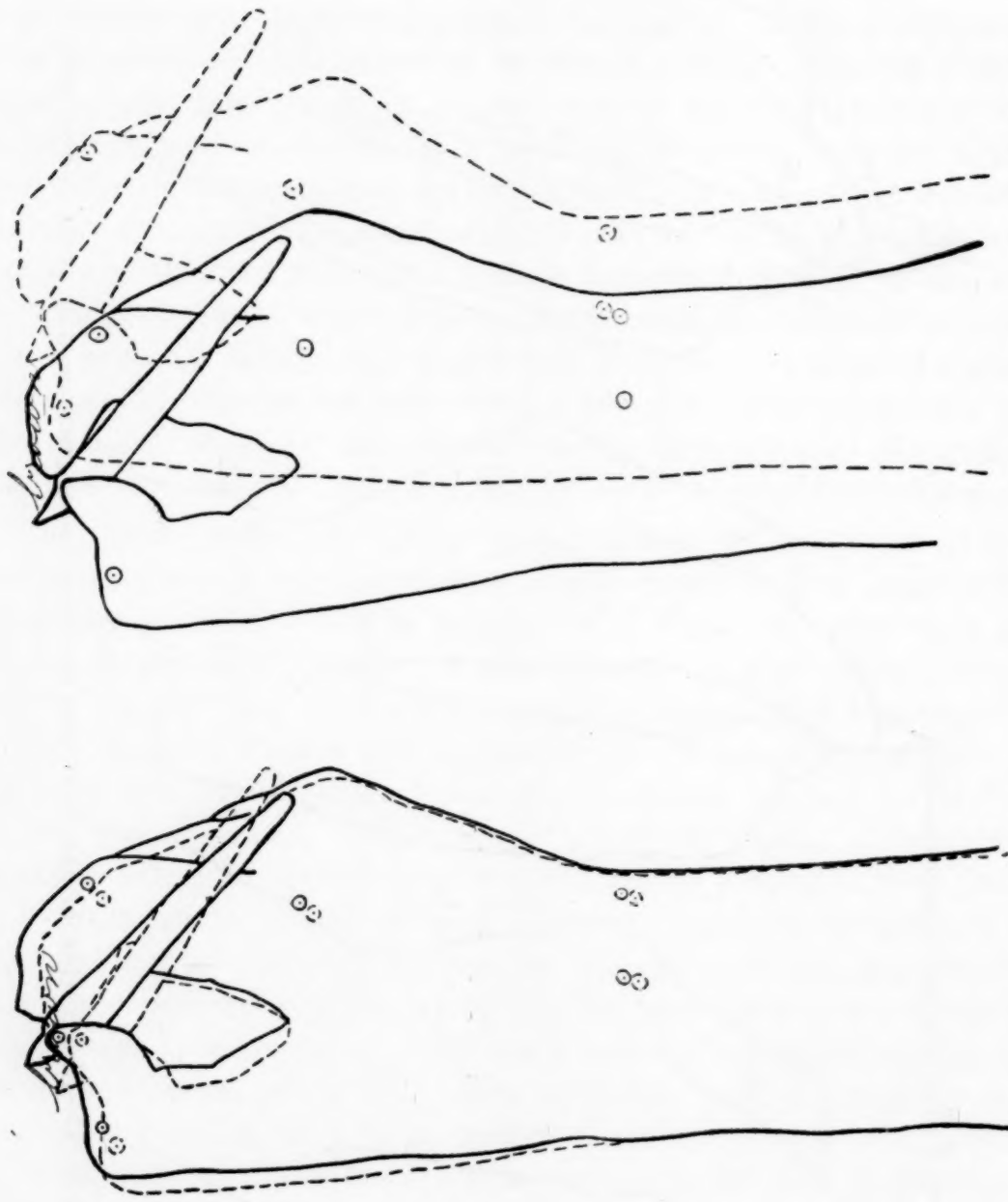

FIG. 2.—Hand position and gross movement of a good writer from Grade VIII

was to classify the individual writers into one of the groups defined in this table. The rest of the table is self-explanatory.

TABLE II.

Basis of Classification of the Individuals According to Their Hand Position.

A. Degree of pronation...	1.	Wrist level, or tilted less than 10°.
	2.	Wrist tilted 10° to 45°.
	3.	Wrist tilted 45° or more.
B. Support of the hand..	1.	On the third and fourth fingers.
	2.	On the side of the hand.
	3.	On the base of the hand.
C. Angle of the arm with the base line of the writing.	1.	Nearly perpendicular.
	2.	10° to 30° from the perpendicular to the right.
	3.	30° to 60° from the perpendicular to the right.
	4.	To the left of the perpendicular.
D. Relation of finger and thumb on penholder.	1.	Finger below thumb.
	2.	Finger opposite thumb.
	3.	Finger above thumb.
E. Looseness of grasp of penholder.	1.	Loose.
	2.	Medium.
	3.	Tight.

Table III indicates, for example, that two individuals of the poor writers in the University Elementary School held the hand so that the wrist was almost level. Ten o fthis group tilted the wrist from 10 degrees to 45 degrees and fourteen tilted it more than 45 degrees. The rest of the table is to be interpreted in a similar manner.

The degree of the pronation of the hand is influenced largely by training. This is shown by the marked change in the distribution of the poor writers in the University Elementary School after training. Whereas only one held the wrist level before training, twelve did so after training. A slightly higher percentage of the public school pupils than of the University children held the wrist level. This reflects the greater amount of formal drill which the public school children had received. Aside from the effect of training, there is little correspondence between the quality of writing and the degree of pronation of the hand. The chief correspondence with quality is found between the two groups of public school children.

Column B represents the manner in which the hand is supported. Here again the effect of training is manifest, both in the comparison of the University Elementary School group before and after training, and in the comparison of these children as a whole with the pub-

lic school children. There is also a slightly greater percentage of the good writers who support the hand upon the fingers than of the poor writers. The poor writers in greater number rest the hand upon the side or upon the base.

TABLE III.

Comparison of the Frequency of Various Hand Positions Among Several Groups of Children.

GROUP	Position	Pronation A No.	Hand Support B No.	Angle of Arm to Base Line C No.	Relative Position of Thumb and Forefinger D No.	Looseness of Grasp E No.
Univ. Elem. School: Poor writers (entire group).	1.....	2	8	5	10	4
	2.....	10	20	20	9	8
	3.....	14	..	2	3	10
	4.....	..	..	2	..	..
	5.....	..	..	..	..	..
Univ. Elem. School: Poor writers who took training.	1.....	1	5	1	8	2
	2.....	9	13	14	4	5
	3.....	8	..	2	2	7
	4.....	..	..	..	..	..
	5.....	..	..	..	..	..
Univ. Elem. School: Good writers.	1.....	2	7	6	9	7
	2.....	4	8	6	2	3
	3.....	6	..	2	..	1
	4.....	..	..	2	..	..
	5.....	..	..	..	..	..
Ray School: Poor writers.	1.....	1	5	2	7	4
	2.....	8	7	6	..	3
	3.....	3	..	2	1	1
	4.....	..	..	2	..	..
	5.....	..	..	..	..	..
Univ. Elem. School: Poor writers after training.	1.....	12	12	7	13	8
	2.....	6	5	13	1	6
	3.....	2	..	1	..	..
	4.....	..	..	..	..	..
	5.....	..	..	..	..	..
Ray School: Good writers.	1.....	4	8	5	8	4
	2.....	7	2	7	..	4
	3.....	1	..	..	..	..
	4.....	..	..	..	..	..
	5.....	..	..	..	..	..

The correspondence between the angle of the arm with the base line and the factors of training and of quality of writing is about the same as in the case of pronation. The effect of training is to reduce the numbers of cases in which the arm is held at an acute angle to the line of writing. Good writers, furthermore, hold the arm nearly perpendicular to the writing in somewhat larger numbers than do poor writers.

The relative position of the thumb and forefinger has a rather marked relation both to training and quality. The effect of training is seen to produce a natural position in which the forefinger rests on the penholder below the thumb. Good writers adopt this position more frequently than poor writers. Finally, a loose grasp is produced by training when it is directed to this aim, as in the case of the University Elementary School group; and in the case of the University children who had had relatively little formal drill the good writers exhibited much more looseness of grasp than the poor ones.

The Composition of the Writing Movement.

Composition refers to the make-up of the total writing coordination. One aspect of it deals with the relation between the arm and finger movement. The facts regarding the more significant phases of composition are presented in Tables IV and V. The interpretation of these tables is similar to that of Tables II and III.

TABLE IV.

Basis of Classification of Individuals According to Their Writing Movement.

A. Movement on letter strokes.	1.	Movement of the whole arm.
	2.	Movement of the hand at wrist joint.
	3.	Movement of fingers mostly at third joints.
	4.	Movement of fingers mostly at second joints (ordinary finger movement).
B. Ease of movement along the line.	1.	Easy.
	2.	Moderately easy.
	3.	Difficult.

TABLE V.

Comparison of the Frequency of Various Kinds of Movement Among Several Groups of Children and Adults.

Group.	Classes of Movement.	Movement on Letter Strokes. A. No.	Ease of Movement Along Line. B. No.
Univ. Elem. School: Poor writers (entire group).	1.....	6	1
	2.....	16	12
	3.....	4	9
	4.....	18	..
	5.....	..	..
Univ. Elem. School: Poor writers who took training.	1.....	3	..
	2.....	12	8
	3.....	2	6
	4.....	13	..
	5.....	..	..
Univ. Elem. School: Good writers.	1.....	3	5
	2.....	9	6
	3.....	3	1
	4.....	8	..
	5.....	..	..
Univ. Elem. School: Poor writers after training.	1.....	11	6
	2.....	6	7
	3.....	1	1
	4.....	2	..
	5.....	..	..
Ray School: Poor writers.	1.....	7	3
	2.....	1	6
	3.....	..	1
	4.....	7	..
	5.....	..	..
Ray School: Good writers.	1.....	5	8
	2.....	5	..
	3.....	..	..
	4.....	8	..
	5.....	..	..
Adults: Poor writers.	1.....	4	..
	2.....	1	..
	3.....	4	..
	4.....	8	..
	5.....	..	..
Adults: Good writers.	1.....	6	..
	2.....	2	..
	3.....	2	..
	4.....	6	..
	5.....	..	..

Column A shows the proportion of children in the various groups who formed the letters by the movement of the whole arm, the movement of the hand, the movement of the fingers produced without bending them at the middle joint, and the ordinary finger movement, which consists chiefly in the hinge movement of the second joints.

We may examine chiefly the frequency of cases which fall under Heads 1 and 4 which comprise the arm movement and finger movement, respectively. There is no evidence in this table that good writers use the arm movement more commonly than poor writers. There is some evidence that training has an effect in producing arm movement, but the effect is not very pronounced. The most interesting fact is that the University group which took training developed a considerable amount of arm movement, in spite of that fact that this was not aimed at directly and was not mentioned in the instructions to the children. These instructions emphasized position, freedom of movement, an easy and fluent sideward movement of the hand, and a rhythmic swing in forming the letters. The training which followed from this type of emphasis apparently produced as an incidental result some arm movement. Undoubtedly the amount of arm movement which is thus directly produced, unlike that which is achieved by being consciously striven for, will remain a permanent element of his writing coordination.

Column B represents a characteristic which it was not easy to measure quantitatively. Pupils were grouped from an observation of the protographs of their writing movement, according as it appeared that they were able to carry the hand smoothly from left to right across the line while the letters were being formed. Ability to do this appears to be a characteristic of good writers as a class and also to be produced by training.

Speed Changes of the Writing Movement.

Probably more important than the position which is assumed in writing is its temporal character. The movement may proceed in a smooth, orderly, comparatively rhythmical fashion, or it may be jerky and irregular. The method of studying these changes has already been mentioned. Figure 3 shows the method in its successive stages. The top line represents the writing itself. The enlarged

reproduction of the word "brown" shows the record on which the position of the pen at the successive exposures of the camera is indicated. The numbers indicate that a succession of exposures took place while the pen remained at a given point. In other words, these represent pauses in the movement. The chart at the bottom shows the speed curve. The vertical units represent the distance which the pen traveled. The horizontal units represent successive twenty-fifths of a second. Each column shows the distance traveled by the pen in a given twenty-fifth of a second. The spaces on the base line at which there are no columns represent pauses. The strokes, which are written underneath the chart, show the parts of the letters which correspond to the part of the speed curve immediately above.

It is evident from an examination, either of the enlarged word or the speed curve, that the writing of this word consisted in a succession of fairly distinct strokes, every stroke separated by a slight pause from the preceding. It is also evident that the pauses in this case come at the natural turning points in the stroke, namely, those points in which the change in direction is marked. The speed on the long strokes is greater than that on the shorter strokes. The speed on a given stroke usually either increases gradually from the beginning to the end or is fairly even throughout the stroke. We have a general picture of regularity and of a correspondence between the length of the stroke and its speed.

This record may be contrasted with that of a poor writer shown in Figure 4. The movement is not divided into units corresponding to the natural units of the letters in the same fashion as in the preceding case. See, for example, the letter *b*. In the first place, there is a pause soon after the beginning of this stroke where we should not expect it. In the second place, the movement is not retarded at the top and the bottom of the long stroke as we should expect. Corresponding with this lack of retardation at the bottom, the stroke is very much rounded. Another illustration may be found in the letters *w* and *n*. There is usually a pause at the end of the last upward stroke of the letter *w*. Here the movement is simply retarded, but not sufficiently to make the sharp turning point, characteristic of this part of the letter. The connecting stroke and the first stroke of the *n* are made into a single stroke by obliterating the characteristic curves at this point. Corresponding to this modi-

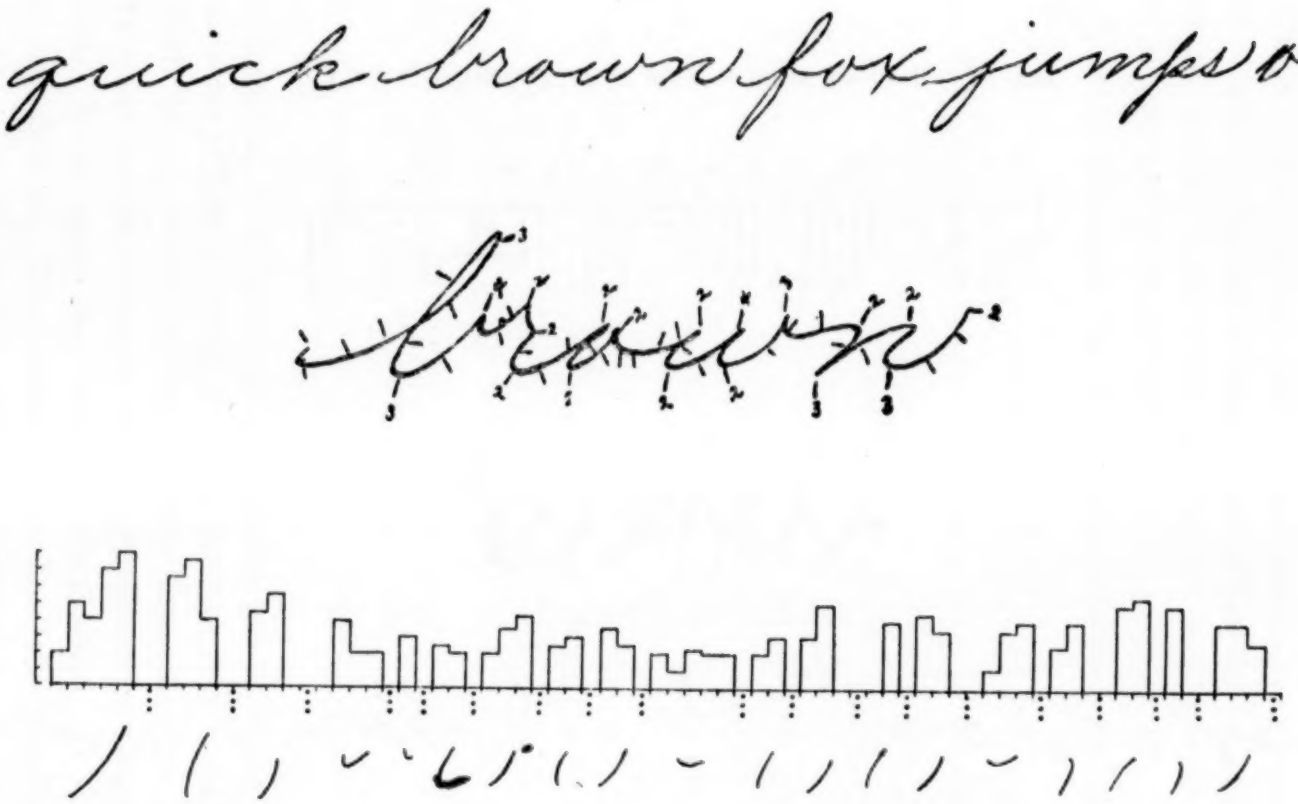

FIG. 3.—Record of a good writer from Grade VIII

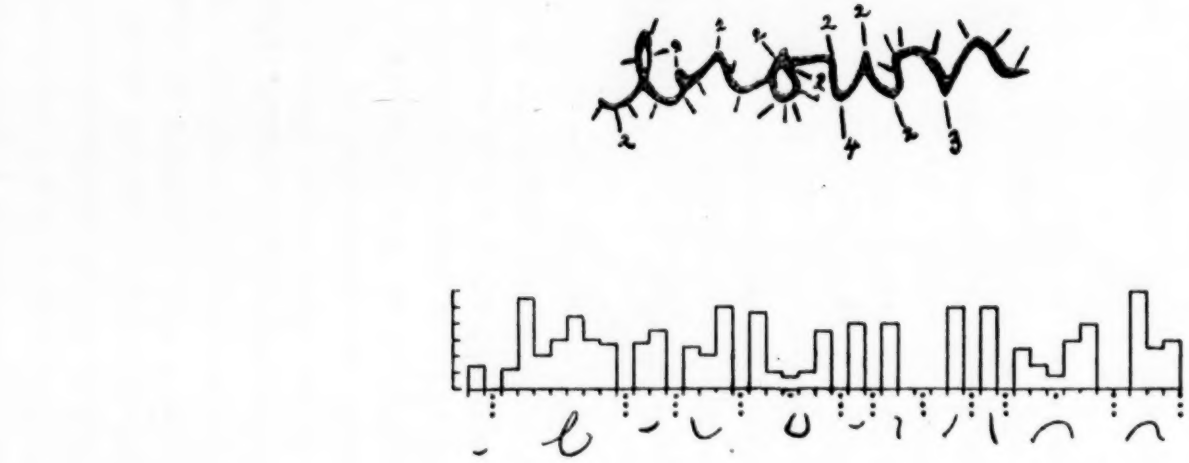

FIG. 4.—Record of a poor writer from Grade VII

fication is the continuity of speed with which this stroke is made. Another characteristic of this writing is that the greatest speed sometimes comes at the beginning or the end of a stroke rather than in the middle. Furthermore, a short stroke, such as the first or the second stroke of the *w*, may be made as rapidly as a much longer stroke. There is also to be found sudden alternation between pauses and rapid strokes. The movement, in other words, is irregular in its succession of strokes and of pauses, and it is not well adapted to the forms which are to be produced.

These illustrations may again be supplemented by the summary Table VI.

TABLE VI.

Percentage of Time Spent by Children in Pauses and at the End of Strokes.

	Grade.	Poorest.	Best.	Before Training.	After Training.
Univ. Elem. School....	VII-B.....	25	27	..	..
	VI-A.....	29	52	..	..
	VI-B.....	28	32	..	..
	V-A.....	16	30	16	34
	V-B.....	32	38	32	49
	IV-A.....	13	36	13	53
	IV-B.....	40	56	40	43
	III-A.....	37	39	37	53
	III-B.....	49	59	49	62
	II-A.....	53	58	53	60
	II-B.....	53	60	53	58
	Average.....	34	44	37	52
Ray School.	VIII-A.....	40	41		
	IV-A.....	32	42		
	Average.....	36	42		

The averages of this table show that the good writer divides the writing movement into a series of units, corresponding to the natural units of form of the letters, more radically than does the poor writer. The differences in the average of the percentages are marked and the difference is the same way in all of the groups. It therefore can hardly be accidental. The older writer and the poor writer both tend to make the successive strokes blend with one another. As practice makes the writer more skillful he is able to bring about a considerable increase in this continuity without the sacrifice of form. When the continuity is carried too far, however, it causes one form to run into another and therefore causes the writing to deteriorate.

SUMMARY OF CONCLUSIONS REGARDING POSITION AND MOVEMENT.

The evidence indicates that the following items of position have some relationship to excellence in writing:

The pronation of the hand to such a degree that the wrist is not tilted more than about 45 degrees from the horizontal;

The position of the forearm at an angle of about 90 degrees with the line of writing;

The support of the hand upon the third and fourth fingers rather than the side or the base of the hand;

A position of the thumb and forefinger on the pen-holder which is produced by slightly bending the joints. In this position the forefinger rests nearer the point than the thumb.

The conclusions regarding speed changes may be summarized as follows:

In good writing the continuity of the movement is broken up sufficiently to produce a retardation or pause at those points where there is a radical change in direction of movement.

The good writer adapts the speed to the strokes by making the longer strokes more rapidly than the shorter ones. The speed on similar strokes is similar.

The organization of the movement within strokes is ordinarily such that the speed is greater in the middle of a stroke and less at the beginning and end.

There is ordinarily not a sharp contrast between the speed of successive parts of the writing, such at would be represented by a sudden, jerky movement.

Experiment With the Training Class.

The susceptibility of some of these characteristics of good writing to training has already been discussed in connection with a group of the University Elementary School pupils. A more extensive experiment of the applicability of such principles, taken as a whole, was made with a school in Kansas City, Kansas, for a period of a year. A system of exercises was worked out in connection with the training class in the University Elementary School. These exercises were mimeographed and taught to the pupils in Kansas City. The progress made by the pupils of this school was compared with the progress of the pupils in two other schools in the same city. The results are summarized in Table VII.

TABLE VII.

Comparison of Average Progress of the Training School (Longfellow) and the Two Check Schools.

Grade and School.	Speed[2]	Form[3]
Grade II:		
Standard	....	...
Longfellow average	29.3	4.0
Bryant average	30.9	2.0
Quindaro average	19.9	2.2
Grade III:		
Standard	13.2	1.4
Longfellow average	25.8	3.4
Bryant average	29.7	—1.2
Quindaro average	18.5	—1.0
Grade IV.		
Standard	7.4	2.0
Longfellow average	28.1	3.2
Bryant average	30.4	—0.7
Quindaro average	38.4	—0.7
Grade V:		
Standard	7.9	1.4
Longfellow average	28.4	3.5
Bryant average	26.7	—0.2
Quindaro average	17.4	0.9
Grade VI:		
Standard	3.7	2.1
Longfellow average	48.6	2.6
Bryant average	11.1	1.5
Quindaro average	—12.1	2.2

[2] In letters per minute.

[3] In units on the Freeman scale.

Grade VII:		
Standard	5.1	1.8
Longfellow average	15.3	7.7
Bryant average	12.3	2.3
Quindaro average	6.4	—1.0
Grade VIII:		
Standard	5.1	1.1
Longfellow average	20.6	6.0
Bryant average	43.8	—0.5
Quindaro average	13.1	0.2
Average Gain:		
Standard	7.1	1.6
Training school	28.0	4.34
Two check schools	20.5	0.43

It is of course impossible to say what features of the method were responsible for the gain which is shown here. The training school gained twenty-eight letters per minute in speed and 4.34 points in form, while the two comparison schools gained 20.5 letters per minute in speed and .43 points in form. The method of producing an improvement in position need not be described. The organization in speed of the writing movement was effected by the application of counting to the writing of the individual letters and of combinations of letters and words. This counting was applied almost solely to the writing of actual letters themselves and not, as is customary, to the writing of exercises. Experience seems to indicate that counting on exercises has a rather remote effect upon the organization of the speed in writing the letters themselves. This investigation, then, confirms, to some degree at least, the conclusions that were drawn from the earlier experimental work regarding the features of position and movement which are particularly important for good writing.

INTELLIGENCE AND ITS MEASUREMENT: A SYMPOSIUM*

XIV. By B. R. Buckingham.

The nature of intelligence and its measurement. Doctors differ in defining intelligence, yet even the man in the street will pass judgment upon the actions of his fellows as either intelligent or stupid. It seems to me that for practical purposes we may get along very well without a definition of the central quality in virtue of which behavior is effective or lacking in effectiveness. This is fortunate because it seems unlikely that any formulation of the nature of this quality will ever be so commanding as to win the assent of a majority even of those who are qualified to pass judgment. We may, it is true, be making virtue of necessity in thus deprecating a description of mentality in the abstract. But even if general intelligence as a personal possession could be defined without reference to its manifestations, the most useful definition would still be in terms of behavior. One who subscribes to this doctrine may occasionally experience a qualm of ingrowing consciousness as he realizes the possibility of thus being dubbed a behaviorist. But perhaps some of us have been called worse names; and aside from the instinctive aversion to being tagged, we shall make shift to endure the epithet with equanimity.

Let us seek, therefore, the nature of intelligence not in some central capacity such as association (Ebbinghaus) or attention (Wundt), but in the nature of acts which the central capacity, whatever it may be, makes possible. The question then becomes one which we can talk about with some chance of being understood. We shall not often nor widely disagree as to whether or not an act in a given situation is effective. Having observed the behavior of a person in a number of similar situations, we rather easily—too easily sometimes—conclude that he has a certain amount of the special ability or aptitude which belongs to the particular type of situation. Again, when we observe a person's behavior, not only with reference to similar situations, but also with reference to a variety of different

---

*Editorial note: We are glad to print this month Dr. Buckingham's statement, which came to hand too late for the April issue. Further interpretation of the material in this symposium will appear in the September issue.

situations, we are justified in concluding that he possesses, to a certain degree in each case, a number of special abilities or aptitudes appropriate to these situations. Finally, with that insistent habit of generalizing which we undoubtedly possess—a habit in virtue of which we seek to reduce diversity to a type, and to express variety in terms of an average—we deduce from the special abilities a general ability. If a person is effective in his behavior in a variety of situations, we assert his general superiority and we ascribe this general superiority to that central directive power which we term intelligence.

It appears that the two essentials in the development of intellectual ability (or the ability to act effectively under given conditions) are, first, a recording mechanism, and, second, something to record. When the late Walter Bagehot characterized the ability of Shakespeare as the product of a first-rate imagination working upon a first-rate experience, he gave, in my judgment, all that was essential.

These, then, are the two variables which enter into both the nature of intelligence and its measurement—the experience and the experiencing nature. If we can control either of these variables, we can measure the other. The experience of school children is reasonably under control. The measurement of the intelligence of school children becomes, therefore, a problem whose solution falls within the bounds of possibility. To my mind, the measurement of the intellilence of persons about whose experience we know little or nothing is highly unreliable. Fortunately, it is with school children that our chief interest lies. The large question which we wish to determine with respect to these children is their educability, or their ability to learn. Whatever may be the theoretical conception of intelligence, it is submitted that it exists for a purpose and must eventuate in something. Pupils who have been brought under the influence of a school regimen will learn somewhat in proportion to their intelligence. A measure, therefore, either of the rate at which learning takes place or of typical products of learning will constitute a measure of intelligence.

Not inappropriately, therefore, do our intelligence tests include tests of learning, such as the digit-symbol test. But by far the greater part of the intelligence tests measure in some way the product of previous learning—i e., they measure knowledge or informa-

tion in a broad sense. This is obviously true, for example, among the verbal tests. It is certain that success in solving analogies depends quite as much upon knowledge of the terms used as upon any innate capacity to perceive relationships. The same may be said of success in picking out synonyms or antonyms, in completing sentences, or in matching proverbs. Success in non-verbal tests likewise depends upon knowledge; or, in other words, upon the product of experience in an experiencing nature. The completion of a picture is impossible unless one knows what the picture ought to be. Similarly, the crossing of a diagram which does not belong with others involves knowledge about the diagrams in question. Even success in following directions requires acquaintance with the terms used and with certain instruments and visual symbols. The non-verbal test, therefore, does not differ essentially from the verbal test except in the kind of the knowledge or experience required to negotiate it.

Accordingly, it seems to me that whatever definition we may give to intelligence in the abstract, we are justified from an educational point of view in regarding it as ability to learn, and as measured by the extent to which learning has taken place or may take place. Of course, in seeking to measure the extent to which learning has already taken place, we do well to avoid specific sorts of learning. Thus, a test of the ability to answer questions in geography would be an inferior measure of intelligence, because it would depend to a large degree upon the amount and recency of a particular type of instruction. There are, however, certain products of intelligence which are of so general a nature that we may appeal to them with some confidence. These we attempt to include in our mental tests in the form of tasks to be performed. The quest for the right kind of tasks is still going on. It is like searching for a few master keys that will unlock many doors.

Next steps in research. I should distinguish at least eight important lines along which investigations should proceed in the measurement of intelligence.

1. The reliability of our present instruments of measurement should be more satisfactorily determined. By reliability I mean the extent to which the scores obtained in the tests are consistent with those obtained by repeating the test with the same individual and under the same conditions (allowing for the effect of practice). When, for example, we infer that a pupil's mental age is ten years,

we wish to know the Probable Error of the result. The public has a right to know the extent to which it may rely upon our determinations.

2. There is at least as great a need for determining the validity of intelligence tests. By validity I mean the extent to which they measure what they purport to measure. If for educational purposes we define intelligence as the ability to learn, the validity of an intelligence test is the extent to which it measures ability to learn. In a very real sense, validity is more important than reliability. No one, for instance, is interested in the consistency of the results of a test which fails to measure the thing it is designed to measure. Such a test would merely be consistently valueless.

3. Most tests consist of a number of sub-tests or tasks each involving a certain type of response. The analysis of the tasks which are now in use and the search for new ones with a view to utilizing a few which are highly symptomatic of intelligence is one of the next steps in research. Here the investigation may proceed along two lines: first, the analysis of intelligence, or, more concretely, of intelligent behavior; and second, the statistical treatment of the results of this analysis through correlation and partial correlation. By the second process the most highly significant of the tasks previously suggested by the analysis of intelligence will be preserved, while at the same time in formulating a battery of tasks duplication will be avoided. The task which correlates most highly with a criterion of general intelligence is one which we wish to use. But if it correlates highly with another task equally related to general intelligence, we shall not use both these tasks in the same battery because each would tell the same story.

4. Just now we are greatly interested in the constancy of the Intelligence Quotient. The bearing of this on prophecy and on long-term planning in school administration is obvious.

5. In addition to instruments measuring general ability, a large number of instruments for measuring specific abilities are desirable. These will have particular value in educational and vocational guidance. A creditable beginning has been made in devising such instruments. "Next steps" should contemplate moving further in this direction.

6. It is generally agreed that intelligence among children and young people differs at different levels of development, not only in

degree, but also in kind. Accordingly, a single mental test cannot be successfully used over a wide range of intellectual levels. As a product of future investigation we hope for a hierarchy of tests, each appropriate to a given level, and each linked with the others by ascertained relationships. Thus for certain purposes we may be able to transmute scores obtained on one test into the scores of another test.

7. It is perhaps a question whether moral qualities and qualities of the "will" should be considered in this connection. It is, however, so clear that success both in school and in life depends upon these qualities, that we cannot avoid pointing out the need which we are under of supplementing our measures of intelligence with measures of the non-intellectual traits which condition human success.

8. Investigators ought seriously to consider the questions of standard procedure and standard form in regard to tests. The simplicity and fullness of directions to examinees, the provision of fore-exercises, the arrangement of material on the printed page, the provision of facilities for checking papers, the directions for handling the scores, and the amount of statistical labor to be required—these are some of the kinds of standardization which we shall doubtless obtain as new tests are devised and the law of the survival of the fittest comes into play.

THE TRUE-FALSE TEST AS A MEASURE OF ACHIEVEMENT IN COLLEGE COURSES

ARTHUR I. GATES,
Teachers College, Columbia University.

Every teacher is painfully aware of the enormous demands upon time and energy required to read and evaluate sufficient examination papers to secure a representative measure of achievement. When classes are many and large, the task becomes a major feature of the "teaching burden" and relief is frequently sought by turnnig over part of the work to "readers" or assistants. That the grading of examination even by teachers of extensive experience is grossly unreliable is indicated by numerous studies carried out within the last decade. The investigations of Cattell, M. F. Meyer, Foster, Finkelstein, Kelley, Starch, Starch and Elliott and others are too well known, however, to the readers of this Journal to necessitate summary here.

In a limited number of fields, reading, writing, spelling, arithmetic, and the fundamental subjects, generally, standardized tests, developed by means of the technique of mental measurements, are available, but in the content subjects of secondary schools and colleges, such tests are few. The unreliability of the conventional essay examination is suspected, but adequate substitutes are wanting.

During the past three years, different forms of examinations patterned after the forms of intelligence and educational tests have been used, together with the conventional essays, and in some cases with tests of intelligence in graduate and undergraduate courses in educational psychology at Teachers College. Sufficient data are now available to make it possible to present a reasonably reliable report of the usefulness of some of these tests.

The Subjects.

Four classes (20, 32, 72, and 74 students, respectively) of an undergraduate course in educational psychology.

Six classes (57, 66, 68, 74, 80, and 82 students, respectively) of graduate courses in intermediate educational psychology. The numbers indicate the total population completing all of the tests upon which the computations are based.

The Tests.

1) Mid-term, or final 1 and 2 hour examinations of the conventional essay type, graded carefully by an instructor, or an assistant, or independently by an assistant and an instructor, or by two instructors.

2) Home-written work; abstracts of books or articles, reports of experiments, essays or special topics, or problems, etc., graded by the instructor, an assistant, or by each, independently. The scores used in correlation are the scores on all such work for a half-semester, representing an average of approximately 8 hours actual writing.

3) The Army Alpha or Thorndike Intelligence Test (Part I or Parts I and II).

4) The new examinations, mostly of the true-false type. The test consists of a series of 30 or more statements which the student checks as true or false, or in some way, such as by underlining, indicates the correct one of two possible answers. The following are sample questions:

A. Questions concerning information given in text or lectures. (Answer "True" or "False.")

1. The Auditory area is in the Temporal Lobe of the cortex.
2. According to Terman's findings, the average intelligence of boys is markedly superior to the intelligence of girls.
3. In learning poetry, the "whole" method is usually found to be superior to the "part" method.

B. Questions concerning information less explicitly given.

1. The curve of forgetting has been more thoroughly investigated than has the curve of learning.
2. If the cerebral hemispheres of a frog were removed, it would remain completely paralysed (assuming it survived the operation).
3. The general laws of learning indicate the wisdom of using "water-wings" in learning to swim.
4. If we find a large percentage of mentally inferior children in the slums, it proves that slum environment is one cause of mental deficiency.

The statements or questions are presented in the following ways: 1) *Oral Presentation.* The instructor reads the question one, two or more times. He may change the wording to make the meaning clear. A little less than one minute per question is sufficient. 2) *Oral and Visual Method.* The instructor reads the question, writes it on the board and reads it again. A trifle more than a minute per question is required. 3) *Visual Method.* Each student is given a mimeographed copy of the questions. A time allowance of one minute per question is sufficient. The time allowances are in all cases generous; in fact, the stress of hurrying is carefully avoided.

Usually the students correct their own results, as the instructor rereads the question and gives the proper answer. The score equals total number of questions minus twice the number wrong. Since this method of scoring is well known and has been defended elsewhere,* it will not be discussed here. Finally the scores are displayed on the board, showing each student his position in the group. A period of 50 minutes is sufficient to give, score, tabulate and discuss the results of a test of 40 questions. Other considerations of the true-false type of examination will be presented in the last section of this paper.

5) The Criterion. In constructing the criterion, the sum of all essay examinations was given the weight of 1.0, the sum of the True-False Tests 1.0, the sum of the written work 0.5; the sum of class recitations, oral quizzes, special conferences, etc., 0.5.

Methods of Computing Results.

All coefficients of correlation presented in the following tables were computed by the Pearson Product-Moment formula.

The several coefficients falling within a decade are presented in a line, thus giving a rough picture of the distributions. The number of cases, the average, the S. D. (σ) of the dist., and the σ (true-obt. av.) are given for each array. In no case has a coefficient been corrected for attenuation for the reason that we are not interested in ideal relationships, but in those which obtain in the existing data. All coefficients listed are positive unless otherwise indicated.

---

*McCall, W. A. Journal of Educational Research, 1920, 1, 33-46.

TABLE I.

Correlations of one essay examination with another.

13, 15,
21, 21, 23, 25, 25, 26, 29
33, 38, 39
42, 47, 49
50, 52, 58
65

N = 19 Average = .35 S. D. = .146 S. D. (tr-obt. av) = .033

TABLE II.

Correlations of one measure of written work with another.

21, 26
32, 33, 35, 37
41, 41, 42, 45, 46
51,
60, 62, 66

N = 15 Average = .43 S. D. = .125 S. D. (tr-obt. av.) = .034

TABLE III.

Correlations of one True-False test with another.

13, 19
21, 24
31, 33, 35, 37, 37, 39
41, 41, 43, 45, 45, 45, 46, 46, 47, 48, 48, 48, 48
50, 50, 51, 53, 53, 53, 55, 55, 55, 55, 56, 57, 58, 58, 58, 59
60, 60, 61, 62, 64, 64, 65, 65, 67, 69
71, 73, 73, 75, 78, 78, 78
80, 81, 81

N = 59 Average = .54 S. D. = .152 S. D. (tr-obt. av.) = .02

TABLE IV.

Correlations of one True-False with one essay examination.

—11
—02
03, 08
11, 11, 13, 15, 18, 19
20, 20, 21, 21, 23, 25, 28
31, 34, 37, 38,
42, 44, 45
51, 57, 59
61, 64, 68
71, 74
84

N = 33 Average = .33 S. D. = .213 S. D. (tr-obt. av.) = .04

TABLE V.

Correlation of one True-False with written work.

00, 04
11, 14, 17
21, 22, 25
30, 33, 37, 38
41
64

N = 14 Average = 25 S. D. = .144 S. D. (tr-obt. av.) = .04

TABLE VI.

Correlations of an essay examination with written work.

15, 23, 26, 31, 53 N = 5 Aver. = .30 S. D. = .12 S. D. (tr-obt. av.) = .05

Inter-correlations of essays, written work and True-False Tests.

The inter-correlations of the essay examinations are decidedly low, for reasons that cannot be discovered from the available data. From other investigations it appears that personal factors involved in the gradings are, in part, responsible. That such may be the case is indicated by the following correlations obtained from the estimates of various sets of essays, each set graded by two judges independently: 27, 37, 41, 46, 50, 52, 66, 74, 90. Average = .54, S. D. = .183. The correlations may be low when the judges accurately estimate different features of the essays which are not positively correlated or they may be low because the judges inaccurately estimate the same features. Many other possibilities make speculations futile. The correlations of estimates of sets of written work, each by two judges, are similarly scattered: 17, 25, 29, 33, 41, 49, 63. Average = .36, S. D. = .144.

The average of the inter-correlations of the True-False tests are higher than those obtained with the essay examinations or the written work; 90 per cent of the coefficients reaching or exceeding the median of the former and 80 per cent reaching or exceeding the median of the latter. It should be recalled that the True-False test represents about 30 minutes' work, whereas the essay requires an hour and the written work about 8 hours. High inter-correlations would not, however, necessarily indicate that the tests are reliable measures of achievement, since they may be produced by many other factors, e. g., capacity to adjust to this particular sort of work. The dispersion of True-False inter-correlations is great. One might speculate about this, or study the conditions under which high and low coefficients were produced (as has been done to some extent), but without illuminating results. Whether the low correlations here found are due to badly selected series of questions, or to low correlations between the different contents tested, each content being accurately tested, to inequalities of previous equipment, to individual variability of performance, or to other causes, is not yet determined. The significance of the inter-correlations will depend upon the correlations of the tests with other tests and the criterion.

A single True-False test gives a correlation with an essay examination which is low, but as high as the average inter-correlation of essay examinations. The correlation of written work with the essay is almost as high, while the correlation of true-false with written work is slightly less. The σ (tr-obt. av.) indicates that real differences between these coefficients are improbable. The True-False test, then, yields as high a correlation with an essay examination or a half year of written work, as would another essay or half year of written work. If it should develop that the True-False tests yield as high a correlation with the criterion as either other test, it would be sensible to use them as *measures of achievement* because of the great saving of time.

TABLE VII.

Correlations of essay examination with criterion of achievement.

27, 28
32, 38
42, 48
51, 51, 53
63, 68
70, 71, 73, 76
82
90

N = 17 Average = .56 S. D. = .187 S. D. (tr-obt. av.) = .045

TABLE VIII.

Correlations of written work with criterion.

13,
25, 27
32, 38, 38
40, 41, 42, 45, 46
53,
61

N = 13 Average = .39 S. D. = .119 S. D. (tr-obt. av.) = .03

TABLE IX.

Correlations of True-False with Criterion.

32, 39
45, 45, 46, 46
50, 51, 51, 52, 52, 52, 53, 56.
61, 63, 64, 64, 65, 67, 67, 68, 68.
71, 73, 73, 74, 75, 75, 77, 77, 78
83, 84, 85, 86, 88
90, 91

N = 39. Average = .65 S. D. = .151 S. D. (tr-obt. av.) = .026

TABLE X.

Correlations of 2 and more True-False Tests with the criterion.
The first row of horizontal figures indicates a class.

	1	2	3	4	5	6	7	8	9	10	Aver.	S. D.
2 T. F. Tests	.46	.52	.56	.59	.65	.69	.72	.82	.84	.86	.672	.127
3	52	59	58	62	69	74	73	84	84	88	.701	.118
4	58	60	63	66	75	79	79	86	88	90	.744	.113
5	68	68	68		83	83	85	86	90	91	.803	.090

TABLE XI.

Correlations of 2 and more essay examinations with the criterion.

	1	2	3	4	5	6	7	8	9	10	Aver.	S. D.
2 essay examinations	40	43	47	49	54	58	60	62	62	66	.541	.085
3 essay examinations	38	52	49	53	56	57	58	65	60	68	.556	.081
4 essay examinations	49		47	49	58	61	58	65	63	66	.573	.068
5 essay examinations			53	60	60	60			64	68	.610	.046

Correlations with the Criterion of Achievement and with Intelligence Ratings.

In constructing the criterion, it will be recalled that the sum of the essays were weighted 1; the sum of the True-False 1, the sum of written exercises, etc., 0.5; class recitations and all other information 0.5.

The correlation of the written work with the criterion is lowest. This may be due in part to the smaller weight given it in the criterion. We are mainly interested in True-False and essay coefficients, and it would appear from Tables V and VI that any weight given the written work would influence these about equally. That the average correlation + 0.65 of True-False with criterion is greater than + 0.56 of essay with criterion is evidenced by the σ (tr-obt. avs.). The difference is .09 $\pm$ σ (tr.-obt. Diff.) .052.

The True-False test thus appears, all things considered, to be the most reliable single measure of achievement. What will be the results when several tests of the same type are combined? How many tests must be given to yield a satisfactory correlation with actual achievement? Tables X and XI give the correlations of 2 to 5 tests with the criterion. The fact that increasing the number of

True-False tests up to 5 brings rather large and uniform increases in the correlations, which is less true of the essay examination, is of marked importance. A group of 3 or 4 True-False tests clearly gives a higher correlation with the criterion than does an equal number of essay examinations.

Just what number of tests is considered "satisfactory" will depend upon our notion of a "satisfactory" correlation and our confidence in the criterion here used. In the course of time a number of cases of sharp disagreement between the scores secured by individuals in the various types of tests have occurred, and of these a number have been further tested as completely as circumstances would permit. The following case may be cited as typical. A student, ranking in the lowest tenth in four True-False examinations, was graded in the upper tenth in an essay examination by an assistant and in the second tenth in another such examination by the instructor. As a result of class-room questioning, two special examinations and a private oral quiz, we were convinced that the True-False results gave the proper measure of achievement. The student was dull but diligent; much material virtually had been memorized by rote. This material, rewritten with excellent penmanship and correct grammar, gave the impression of general mastery of the subject. Ability to apply the facts and principles was unusually deficient. Since nearly all of the special cases turned out similarly, the implication is that the better the criterion of achievement, the greater the predictive value of True-False as compared to the essay. Experience with the True-False examination has convinced us that it affords an opportunity to test more adequately the power to apply and use information in solving a concrete problem or meeting a novel situation.

The correlations with intelligence ratings will be of interest in this connection. Tables XII, XIII and XIV show average correlations with intelligence as follows: A single True-False + .406, an essay + .344 and written work + .255. The difference between True-False and essay is + .062 which is about the same as the σ (tr.-obt. diff.) $= \pm .067$. From tables XV, XVI and XVII it appears that combining several essay examinations makes no noticeable increase in the correlation, whereas the addition of True-False tests yields a steadily increasing correlation. A team of 3 or 4 True-False tests gives a clearly higher coefficient than a similar team of essays.

TABLE XII.

Correlations of Written Work, with Intelligence Ratings.

08
17, 18
22, 25, 26
34, 39
41

N = 9 Average = .255 S. D. = .102 S. D. (tr.-obt. av.) = .034

TABLE XIII.

Correlations of essay exam. with Intelligence.

06
13, 15, 19, 19
25, 29, 29
32, 35, 36, 37
41, 47
52, 56
61
72

N = 18 Average = .344 S. D. = .170 S. D. (tr-obt. av.) = .04

TABLE XIV.

Correlations of True-False with Intelligence.

13, 15
21, 24, 29, 29
32, 36, 39, 39, 39
41, 45, 45, 46, 47, 49
59, 56
60, 66,
71

N = 22 Average = .406 S. D. = .15 S. D. (tr-obt. av.) = .032

TABLE XV.

Correlation of the two Half-Term measures of written work with Intelligence, for four classes.

17, 24, 30, 38, Average = .272

TABLE XVI.

Correlation of 2 or more essay examinations with Intelligence, arranged by classes.

	1	2	3	4	5	Aver.	S. D.
2 essays	24	29	35	39	43	.34	.052
3 essays	26	28	39	39	40	.344	.057
4 essays	30	29		40	41	.350	.054

TABLE XVII.

Correlations of 2 or more True-False with Intelligence arranged by classes.

						Aver.	S. D.
2 True-False	33	38	45	49	51	.432	.067
3 True-False	38	44	43	54	57	.472	.071
4 True-False	41	47	52	59	55	.508	.063
5 True-False	45	52		63	58	.545	.063

That the True-False test is not merely a new test of general intelligence is indicated by the fact that it yields higher correlations with the criterion of achievement; correlations which are higher than those obtained by the conventional examination. It would be useful to compare the correlations yielded with intelligence and with achievement by True-False tests composed of questions of information with questions requiring applications or problem solution, but it cannot be readily accomplished with the available data.

Correlations Obtained by Different Methods of Presentation.

The correlations of single tests with the criterion given in Table IX have been classified according to the method of presentation with the following results:

1) Oral presentation, n = 18, average = .65 σ (tr.-obt. av.) $\pm$ .043.
2) Oral and visual (written on board), n. = 8, average = .62, σ (tr.-obt. av.) = .048.
3) Visual (mimeographed), n. = 13, average = .66, σ (tr.-obt. av.) = $\pm$ .030.

While the data are too few to make generalization secure, no method of presentation shows superiority.

Summary.

The value of a True-False examination as a measure of achievement in college courses is demonstrated by the following comparisons:

1. It yields correlations with other tests of the same type, averaging + .54, σ (tr.-obt. av.) $\pm$ .02, as compared to inter-correlations of essay examinations of + .35, σ (tr.-obt. av.) $\pm$ .033, and inter-correlations of written work of + .43, σ (tr.-obt. av.) $\pm$ .034.

2. The True-False tests yield a correlation of + .33; σ (tr.-obt. av.) $\pm$ .04 with an essay examination, which is as high as inter-correlations of essay examination.

3. The True-False test yields a correlation with written work of + .25, σ (tr.-obt. av.) $\pm$ .04, which is probably as high as the correlation of essay examination with written work, + .30, σ (tr.-obt. av.) $\pm$ .05.

4. The correlation of a True-False test with the criterion of achievement is + .65, σ (tr.-obt. av.) ± .026, as compared to + .56, σ (tr.-obt. av.) ± .045, for the essay examination and + .39, σ (tr.-obt. av.) ± .03 for the written work.

5. Combining several True-False tests produces markedly higher correlations with the criterion which is not true of the essay examinations.

6. Correlations with intelligence tests are: for the True-False + .406, σ (tr.-obt. av.) ± .032, for the essay examinations, + .344, σ (tr.-obt. av.) ± .04 for the written work, + .255, σ (tr.-obt. av.) ± .034.

7. Combining several True-False tests produces a marked increase in the correlations with intelligence, which is not true of the essay examinations.

8. Intensive study of certain cases indicates that the better the criterion of achivement, the greater the predictive value of True-False as compared to the essay examination.

Other Advantages of the True-False Examinations.

The following advantages of the True-False examination are not derived from objective data, but are accumulations of opinions generally voiced by those who have given or those who have taken such tests.

1. It saves an enormous amount of time. All the work of scoring is saved, since it is instructive to have the students correct the papers.

2. It is possible to develop standards of achievement by which one class may be compared with others.

3. Since each test includes 30 or more questions, it is possible to examine the field much more thoroughly than with the conventional examinations.

4. More refined units of measurement are secured; the distribution is wider than can be reliably obtained by the conventional examination.

5. Students report that the True-False examination is conductive to more effective methods of study. They study to understand and apply, rather than to commit to memory for purposes of sheer reproduction.

6. The True-False examination is really an excellent teaching instrument. By correcting the answers at once, haziness, misunderstandings and ignorance of facts and principles are clearly shown.

7. Showing the student in position in the group by distributing the scores on the blackboard is of recognized pedagogical value.

8. According to votes, at least 90 per cent of the students express a preference for this type of test. Some of the reasons frequently given are:

a) When the hour is over, the results are known; there are no days of uncertainty to be suffered.

b) After the initial adjustment, less nervousness before and during the examination is felt, usually.

c) Anticipation of being misunderstood, of favoritism or the opposite, is unnecessary.

d) The work during the examination is less exhausting. There is little writing or eye strain; no need of fear of being unable to finish within the time limit.

TRANSMUTATION OF VALUES ON THE THORNDIKE AND AYRES HANDWRITING SCALES: A CORRECTION

T. L. KELLEY,
Leland Stanford Junior University.

My attention has recently been called to a numerical error which occurred in a study of mine appearing in THE JOURNAL OF EDUCATIONAL PSYCHOLOGY of December, 1914, entitled "Comparable Measures." I take this opportunity to resubmit conclusions based upon correct calculations. This article dealt with the comparability of scores upon the Ayres and Thorndike Handwriting Scales, and gave an equating of the two scores, which is in error. The constant 2.917 reported on page 593 should be 2.223. The constant 7.90 derived from the inaccurate value 2.917 should be 7.147. Other values derived from these are in error. I list below the materially inaccurate statements and following them the revised correct statements.

Inaccurate statement: "The average variation of the estimate (of handwriting) upon the Ayres scale is 7.40 and upon the Thorndike scale when reduced to comparable units 8.68, which gives a difference of 1.28 in favor of the Ayres scale. The probable error of this difference is unknown, but is greater than .372."

Correct statement: "The average variation of the estimate (of handwriting) upon the Ayres scale is 7.40 and upon the Thorndike scale when reduced to comparable units 7.85, which gives a difference of .44 in favor of the Ayres scale. The probable error of this difference is unknown, but is greater than .372."

Incorrect Statement. Equal Measures.		Correct Statement. Equal Measures.	
Ayres Scale.	Thorndike Scale.	Ayres Scale.	Thorndike Scale.
9.5	5	13.3	5
17.4	6	20	5.96
20	6.33	20.5	6
25.3	7	27.6	7
30	7.60	30	7.34
33.2	8	34.7	8
40	8.86	40	8.74
41.1	9	41.9	9
49.0	10	49.0	10
50	10.13	50	10.13
56.9	11	56.2	11
60	11.39	60	11.53
64.8	12	63.3	12
70	12.66	70	12.93
72.7	13	70.5	13
80	13.93	77.6	14
80.6	14	80	14.33
88.5	15	84.8	15
90	15.19	90	15.73
96.4	16	91.9	16

DEPARTMENT FOR DISCUSSION OF RESEARCH PROBLEMS

Conducted by LAURA ZIRBES

This department has a two-fold function. It aims to serve research workers as well as educators, whose work brings them in close contact with children in the schools. It hopes to accomplish this service by suggesting research studies, which will meet well-defined school needs.

In order that this service may be real and effective, the co-operation of research workers and school people is desired. Correspondence with reference to the following qustions will be considered in selecting topics for future discussions.

a. Which of the studies proposed would help you to solve a practical problem?

b. What topics might well be added to this list? Replies may be addressed to: Miss Laura Zirbes, 646 Park Ave., New York City.

SOME THINGS I WANT TO DO OR SEE OTHERS DO

WM. A. McCALL,
Teachers College, Columbia University.

1. *Technique for Constructing Validation Criteria.* There are many steps in the process of scale construction. Among these the following may be mentioned: (a) Preparation and organization of test material, (b) Preparation of instructions for applying and scoring the test, (c) Validation of test, (d) Scaling the test, (e) Determination of reliability, objectivity, norms, and perhaps standards.

Of these five steps one is usually omitted in whole or in part, and more often *in whole* than *in part*. Unfortunately the omitted step—validation of the test—is the most important step, for though a test be perfect in every other respect, it is worthless and may be actually harmful if it does not measure what it purports to measure.

Test validation has not been omitted because test constructors have failed to realize its importance, nor has it been omitted because test constructors are lazy. A test can be validated only by demonstrating that it yields a substantial positive correlation with some objecively defined criterion outside of the test which is being validated. Test validation has been omitted frequently because of the great difficulty of objectively defining the criterion. It may be possible to evolve a general procedure and it is certainly possible to

formulate specific procedures for the construction of criteria for evaluating not only mental tests, but also other measurements of individuals. Here is a fit task for both genius and industry.

2. *Area, Volume, or Weight of a Mental Function.* The size of a man is not measured by his height, or width, or thickness, but by a certain balanced combination of these aspects of size. A man's volume is the real measure of his size, for volume gives to every dimension of size its proper influence. Because of a reasonable uniformity in the contents of men, weight is a fairly satisfactory combiner of a man's size dimensions.

Mental traits have area or volume very much as a man has size. There was once a time when no method had been evolved for determining the volume of physical objects. Mental measurement is in exactly that stage of evolution at present. Reading ability, for example, has three dimensions at least—accuracy, speed, and difficulty of the passage read. Furthermore, that experimenter is lucky or obtuse who has not been forced to worry about how to weight time, or errors, or other aspects of mental performance.

The common current method of determining the volume of a mental function is a purely impressionistic one. Estimations of the volumes of objects are risky. Subjective judgments of the volumes of mental traits are truly hazardous. Even the most competent judge has little confidence in his estimate.

I suggest the following methods for determining the volume of a mental function, and express the hope that some of us will find the opportunity to determine the correspondence between them, evaluate each, add others, and finally construct a formula for computing the volume of the mental functions measured by the more widely used standard tests.

(a) Subjective estimate. Has the pupil who reads with a speed of 20, an accuracy of 20, and a difficulty of 20 more reading ability than one who reads with a speed of 25, an accuracy of 15, and a difficulty of 21? (b) Subjective estimate of social worth. It might, for example, be argued that while high speed and low accuracy are as brilliant an illustration of neutral behavior as correspondingly lower speed and greater accuracy, still, the latter should be given a higher score because of the relatively greater social value of accuracy. (c) Weight, i. e., some one-dimensional measurement which represents a condensation of speed, accuracy, and difficulty just as

the weight of a man is a condensation of his size dimensions. (d) Correlation with some criterion which is otherwise defined and objectively measured in some one-dimensional manner. Once the criterion is secured this can be done by means of partial correlation and a regression equation. Once given an intelligence criterion, for example, it is a simple matter to determine the optimum weightings for rights, errors, speed, etc., in a proposed intelligence test. (e) Experimental determination of the equality of various combinations of speed, accuracy and difficulty. Courtis is probably the only one who has made an extensive use of this method. (f) Self-correlation between two trials of equivalent tests upon identical pupils. This method assumes that the proper weighting for each of speed, accuracy, and difficulty in the composite of the three is that weighting which will yield the highest self-correlation between the two composites. There is a statistical procedure which will facilitate the determination of these weightings. The self-correlation method has never been used for determining the volume of a mental trait.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

Equality in Difficulty of Alternative Intelligence Examinations, Edward L. Thorndike, Journal of Applied Psychology, 1920, 4, 283-289, suggests a method of constructing alternative forms of tests and gives data secured with different forms of the Thorndike Intelligence Examination.

Correlation of Army Alpha Intelligence Test, with Academic Grades in High Schools and Military Academies. H. E. Burtt and G. F. Ayrs. Journal of Applied Psychology, 1920, 4, 289-294. Finds a correlation of about + 0.40.

A Comparative Study of the Intelligence of Seventy-five White and Fifty-five Colored College Students by the Stanford Revision of the Binet-Simon Scale. S. M. Derick. Journal of Applied Psychology, 1920, 4, 316-330. Whites yield an average I. Q. about 9 points higher than negroes.

The Selection of Mill Workers by Mental Tests. Arthur S. Otis. Journal of Applied Psychology, 1920, 4, 339-342. A high degree of intelligence a detriment for certain tasks.

Intelligence Tests and Academic Standing. Edna Gordon and H. J. Baker. Journal of Applied Psychology, 1920, 4, 361-363. Correlations range from 0 to + .54, depending on the course of study.

Picture Completion. Robert H. Gault. Journal of Applied Psychology, 1920, 4, 310-316. Suggestions concerning the weighting of responses in the Healy test.

Reconstruction of Mental Tests. Beardsley Ruml. Journal of Philosophy, 1921, 28, 181-185. A reply to an article by Pressey on statistical methods.

A Comparison of Two Methods of Giving the Number Series Completion Test. John E. Anderson. Journal of Applied Psychology, 1920, 4, 346-348.

Education of Juvenile Delinquents. Edgar A. Doll. Journal of Delinquency, 1921, 6, 331-347. A discussion of methods of diagnosis and educational treatment.

The Importance of Physical and Mental Examinations as an aid to Treatment and Training in a Reform Institution. Edmund B. Hilliard. Journal of Delinquency, 1921, 6, 347-355.

Irregularity in Intelligence Tests of Delinquents. Julia Mathews. Journal of Delinquency, 1921, 6, 355-362. Data presented suggests the need of a more careful inquiry into the causes of "scattering."

An English Form Test. Thomas B. Briggs. Teachers College Record, 1921, 22, 1-11. Describes two equivalent forms of a test for formal English to be used in grades VII, VIII and IX. Each test consists of 20 short passages to be punctuated by the subject.

Proposed Uniform Method of Scale Construction. **William A. McCall. Teachers College Record, 1921, 22, 31-51. Describes a method of scale construction in which the mean performance of children between 12.0 and 13.0 years of age is taken as the reference point, with 10 S.D. as the range and —5.S.D. as the zero point. The range is divided into 100 units, which serve as the scale points. A reading scale constructed by the use of this method is described.**

A Qualitative Investigation of the Effect of Mode of Presentation Upon the Process of Learning. **F. J. O'Brien. American Journal of Psychology, 1921. 32, 249-284. Suggests the need of supplementary studies of learning by introspective accounts of imagery employed.**

The Child Mind. **H. J. Mulford. American Journal of Psychology, 1921, 32, 179-196. A variety of opinions on the child's mind from a genetic point of view.**

On the Relevancy of Imagery to the Process of Thought. **C. Comstock. American Journal of Psychology, 1921, 32, 196-231. A discussion of several functions of imagery in solving problems.**

The Use of a Time-Record Blank in the Standardization and Supervision of Student-Teaching Courses. **William S. Gray. Educational Administration and Supervision, 1921, 7, 121-133.**

The Training of Teachers to Supervise Study. **Alfred L. Hall-Quest. Educational Administration and Supervision, 1921, 7, 160-166.**

Some Second Lieutenant Psychology in School Administration. **Garry C. Myers. Educational Administration and Supervision, 1921, 7, 171-175.**

A Measure of Ability to Judge Poetry. **Allan Abbott and M. R. Trabue. Teachers College Record, 1921, 22, 101-127. Twenty-six sets of 4 poems each, graded by competent judges, and scaled in steps of merit.**

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *Psychology for Normal Schools.*—This text book[1] for normal school students is novel in its organization. Instead of the conventional chapters, the material is arranged in 46 lessons, averaging 7 pages each, introduced by a series of suggestions for observational work and followed by "Topics for Special Study and Report," "The Lesson Applied" and "Selected References."

One hundred and twenty pages are devoted to instinctive behavior, and nearly 50 pages are devoted to the general studies of heredity; then follow 20 pages on habit; 65 pages on sensation, perception and other mental progresses, and a final 60 pages on individual differences. Although the divisions of the book suggest a dynamic psychology, the material is almost wholly descriptive. Instincts are named and described in terms of children's familiar play and other activities. The sections on habit, learning and the mental processes—sensation, perception, attention, imagery, memory, imagination, thinking and will—consist primarily of definitions and illustrations. The laws of learning are nowhere stated; no reference is made to the experimental work upon the acquisition of skill or information, upon problem solving or thinking. No mention is made of a curve of learning or forgetting, of economical methods of memorizing or the psychology of the school subjects; the words spelling, arithmetic and writing do not appear in the index. The chapter on "Memory and Imigination" is typical; memory is defined as "reproductive imagination." Three pages are used to amplify this, and the remainder of the short chapter is devoted to "Earliest memories of childhood" and "Desultory memory of childhood."

The experimental psychology of individual differences is condensed into about 40 pages. Many of these are devoted to definitions of "normal", "subnormal", "epilepsy", "hysteria", etc., but no graph or mention of the curve of distribution is to be found. A half page is

---

[1] *Psychology for Normal Schools,* Lawrence A. Averill, Boston; Houghton Mifflin Company, 1921, Pp. XX + 362.

devoted to the discussion of educational tests; two-thirds of a page to group intelligence tests.

The author, in the preface, finds fault with current texts for normal schools because of "their scientific terminology and style of language," which he has sought to remedy. Whether the cure is better than the malady, the reader may judge from the following sample effort to explain "introspection" as a method of investigation in psychology. The teacher had told a story of primitive firearms which aroused a pupil to exclaim, "My father's got one of them guns up in our shed-chamber." To this the teacher retorted, "How often must I tell you, James, to be careful of your English" The author adds: "Now, you will understand better why introspective cleverness is a virtue on the part of a teacher. James' teacher was not an introspectionist; she failed to see behind James' ungrammatical response the real impulsion that motivated him; she saw not the thing that James saw, but only the poor language in which it was coached."

A. I. G.

---

2. *Psychology for Nurses.*—The popular tendency to apply the term psychology to any suggestion or opinion concerning the behavior of human beings is exemplified in this book.[2] A great deal of advice for nurses and many observations of people generally are grouped under the headings of textbooks in psychology. Occasionally a scientific fact or principle is stated and illustrated by hospital experience, but most of the psychology consists of definitions quoted largely from Thorndike's "Elements" and James' "Briefer Course," which appear in an appendix to each chapter. The only other authors mentioned are Pillsbury, Colvin and Bagley, and in one place Terman. The following quotation, picked at random, from the chapter on "Sensation," illustrates the procedure.

V. Dermal Sensations.

The Nurse's Use of Sensations of the Skin. Sense-organs of the skin have much to do with our responses. The nurse, like the surgeon, needs "eyes in her fingers." The nurse gets all manner of information by way of touch. It

[2]*The Psychology of Nursing.* Aileen Cleveland Higgins, G. P. Putnam's Sons: New York, 1921. Pp. XVI + 337.

is useful to be sure of one's hands in an emergency, particularly when they must act as guides in darkness or half-light. To gauge accurately the amount of pressure necessary in various nursing procedures is fundamental in good technique.

3. *Psychology for Musicians.*—The first 60 pages of this small book[3] give tabulations and discussions of the results of a questionnaire study in which 100 musicians and composers gave replies to a series of 24 questions with regard to methods of learning and teaching, and the effects of various factors on performance. As is usual in such studies, it is difficult to judge what some of the questions and many of the answers mean. It appears that no one type of imagery is employed in learning and performing, although visual, auditory and kinaesthetic representations are most (and about equally) frequently employed. In learning "it is not the sensory field that aids so much as it is the purely intellectual processes." Sixty-six musicians learn "bit by bit," 14 by the "whole method" and 14 by "various combinations of the two." Many assert that emotions (kinds not specified) assist to make performance successful, but add that the emotion must be kept under control. Twenty-one play better when alone, while 65 are favorably stimulated by a sympathetic audience. Nervousness, anger, excitement or disturbances in the audience generally reduce efficiency. Most musicians believe that native aptitude rather than any special form of practice is responsible for their success.

The second part of the book contains a rather superficial exposition of the principles of psychology, consisting chiefly of a series of descriptive definitions of the conventional mental processes. A. I. G.

---

4. *The Work of a Juvenile Court.*—Franklin C. Hoyt,[4] Judge of the Children's Court of New York City, has described the principles and machinery upon which the practice of dealing with juvenile offenders is based in a series of narratives designed for the popular reader. The book "does not pretend to cover any particular phase of child psychology, nor is it written with the slightest idea of serving as a manual on juvenile-court work in general." Since it de-

[3] *The Musician's Mind.* Antoinette Feleky, New York: Pioneer Publishing Company. 1921. Pp. 108.

[4] *Quicksands of Youth.* Franklin Chase Hoyt, New York: Scribners, 1921. Pp. XI + 241.

scribes at length the action of the court in a series of real cases, it will be of interest to many students of education and psychology.

A. I. G.

---

5. *Two Books Dealing With The Principles of Education.*—Taking as a thesis "the claim of individuality to be regarded as the supreme educational end," Professor Nunn has written a stimulating survey[5a] of the whole field of educational theory and practice. The author is throughout alive to scientific research, movements as recent as the use of intelligence tests for admission to Columbia College being considered. The book, in fact, gives an interesting interpretation of current scientific work. In discussing the nature of development, the writings of Pearson, Spearman, Binet, Terman, and especially those of McDougall and Shand, Hart, Freud, Adler, Prince and others are mentioned. The technical studies of Book, Huey, Ach, and Burt are considered. The author marshals experimental findings and current opinion to support his theory that the main function of the school is to socialize its pupils, but points out in a final chapter that such a function "in no wise contradicts the view (stated at the beginning) that its true aim is to cultivate individuality."

Professor Coursault's book,[5b] while leading to a somewhat similar conclusion, has a quite different approach. It is based primarily upon theoretical writings in education rather than upon scientific investigations and theories in related fields. The influence of Dewey, Royce, MacVannel, Bagley and McMurry appears clearly. The author attempts to harmonize the views of man as a psycho-physical organism in a process of adjustment through stimuli and responses with the tleeological view of man as a person controlled by purposes and ideas. The importance in the educative process of the appreciation of values and the development of purposes as compared to the importance of information is stressed. In explaining how purposes are developed and means of control are made, the author analyzes, in a fashion similar to Dewey, the processes of thinking. In this connection the nature and function of history, literature and

---

[5a] *Education, Its Datan and First Principles.* T. Percy Nunn, New York: Longmans, Green and Company, 1920. Pp. VII + 224.

[5b] *The Principles of Education.* Jesse H. Coursault, Boston: Silver, Burdett and Company, 1920. Pp. XII + 468.

other fine arts and the methods with which subject-matter should be taught are explained. The book will serve as a useful introduction to current educational theory and practice. A. I. G.

---

6. *A Practical Textbook Interpreting Biological Psychology.*[*]— Rarely do we find a textbook on psychology which is designed to meet the needs of teachers in service and those preparing to teach. If this text had the word "Psychology" in its title, the author might have been accused of using the word, in view of the fact that the contents are not "pure" psychology. Any text which is to lead teachers to make use of the findings of research in the solution of their problems must seek to present its materials in such form that the practical bearings of the principles of psychology are made apparent in illustrative instances. It must also provide some practice in the solution of such problems if instruction is to function in practice.

At the risk of appearing too discursive, this text has interpreted those fundamental conditions governing mental development and shown their relation to the problems of the classroom and the school. In this the author has made his text conform to the psychological requirements of a limited field, avoiding technical and speculative discussions which might make for a more systematic and comprehensive but a less dynamic treatise.

The book is written around two questions: 1. How does the individual normally respond at different periods in his development to the typical situations—physical, intellectual, esthetic and social—in which he is placed? 2. How can he best appropriate the materials and benefits of education so that he can utilize them to greatest advantage in daily life? Part I deals with the dynamic aspects of mental development. Part II restates these and gives educational interpretations. Part III provides exercises in analysis, interpretation, investigation and application. This problem material is full of quotations from other writers, all carefully correlated with the preceding chapters of the book. It occupies over one hundred pages, and furnishes a wealth of opportunities for practice in the sort of thinking which is reflected in improved teaching. L. Z.

---

[*]M. V. O'Shea. *Mental Development and Education.* The MacMillan Co., New York, 1921. VII 403.

7. *A Study in the Correction of Speech Defects.*[7]—This pamphlet reports the analytical study of the speech disorders of over one hundred children, with a classification of defects and their causes. Two type cases are reported in great detail, with accompanying graphic records of the effect of remedial training and a controlled regime. These case studies are perhaps the most valuable and illuminating part of the contribution. They are followed by materials for speech examination and individual speech records, which make it possible to measure improvement objectively. By making pupils aware of their progress in overcoming defects such records may become a psychologically potent remedial factor. Other remedial suggestions are given, together with a bibliography. L. Z.

---

8. *A Report on Teacher Training Departments in High Schools.*[8] Minnesota has attempted to improve the quality of instruction in her rural schools by offering normal training courses in the high schools of small towns. The movement began about twenty-five years ago in a very small way, and was brought to the efficiency described in the report, under the direction of Miss Mabel Carney, whose program included special college or normal training for the instructors, uniform and definite courses throughout the State, and practice teach, ing in rural schools. The movement received an impetus when increased State aid made more balanced courses possible, and when the completion of the prescribed courses resultd in exemption from teachers' examinations. The plan does not pretend to be the ultimate solution of the problem of providing trained teachers for rural schools. It has been a successful temporary expedient. The program toward which the State Department in Minnesota is working contemplates equal training for city and rural teachers and close affiliation with normal schools. In the meantime, the training departments in high schools are being improved to meet the immediate needs more satisfactorily. The appendix contains (1) a comparative table showing facts concerning training departments in 13 other States, (2) a suggestive yearly scheme for practice teaching, (3) reports, (4) a bibliography. L. Z.

---

[7]Sarah M. Stinchfield. *A Preliminary Study in Corrective Speech.* Studies in Child Welfare, Vol. I, No. 3. University of Iowa, 1920. Pp. 36.

[8]Lotus D. Coffman. *Teacher Training Departments in Minnesota High Schools.* General Education Board. New York, 1920. Pp. VIII + 92.

9. *Report of the Proceedings of Schoolmen's Week.*[9]—Because the contents of this yearbook duplicate the variety and scope of the program of Schoolmen's Week, it is impossible to give any adequate idea of the material in this limited space. While there is a preponderance of administrative problems, there are only two discussions dealing definitely with curriculum reconstruction. Journal readers will be especially interested in the report of the Bureau of Educational Measurements and in a number of articles dealing with intelligence tests and their uses. L. Z.

---

10. *The Educational and Intellectual Status of the Army Medical Service.*[10]—Records for approximately 2500 medical officers were available for the statistical investigation reported in this bulletin. The assembled data justify the statement that "the Medical Corps obtained the services of the ablest as well as the weakest men of the profession." The general intelligence rating of medical officers is lower than that of other branches of the service, with the exception of the Dental and Veterinary Corps, and is practically the same as that of the Quartermasters' Corps. But when the psychographs or curves representing measurements for each of eight types of tests are studied, typical differences between those of the medical corps and other arms of the service noted. Those of the several medical groups are similar in certain respects. The median length of schooling was 15.8 years, and 11.07 was the median in years of experience. There is a remarkable difference in the frequency of superior and very superior intelligence in the Medical Corps as compared with the Engineer Corps, and the difference is distinctly in favor of the Engineers, as is shown by tables and graphs. The most significant tabulations and graphs are those in which the achievement in each of the eight parts of the intelligence tests are set down for the various arms of the service, and the various branches of the medical service.

Although the admission of men of inferior ability was due largely to the emergency, it remains true that these men would have practiced medicine even though barred from the army by more rigid re-

---

[9]*Seventh Annual Schoolmen's Week Proceedings*: University of Pennsylvania Bulletin. Vol. XXI, No. 1. Philadelphia, 1920. Pp. 336

[10]M. V. Cobb and R. M. Yerkes. *Intellectual and Educational Status of the Medical Profession as Represented in the United States Army.* Bulletin of the National Research Council. Vol. I, Part 8, No. 8. Washington, D. C. 1921. Pp. 75.

quirements. This study inclines one to the belief that intelligence requirements should be set up for admission to the medical schools and to the profession. A calling so closely related to the national well-being needs the best minds as well as the best training. L. Z.

III. Additional Publications Received.

A. Books in General and Applied Psychology.

Paton, Stewart. *Human Behavior.* New York: Charles Scribner's Sons, 1921. Pp. V + 465.

Tracy, F. *Psychology of Adolescence.* New York: MacMillan Company, 1921. Pp. X + 246.

B. Mental and Educational Tests.

Courtis, S. A., and Shaw, L. A. *Courtis Standard Practice Tests in Handwriting.* New York: World Book Co., 1921.

Chapman, J. Crosby. *Trade Tests.* New York: Henry Holt & Co., 1921. Pp. IX + 435.

Trabue, M. R., and Stockbridge, F. P. *Measure Your Mind.* Doubleday, Page & Co., 1921. Pp. VII + 349.

C. Publications in the General Educational Field.

Finney, R. L. *The American Public School.* New York: MacMillan Company, 1921. Pp. XIV + 335.

O'Brien, J. A. *Silent Reading.* New York: MacMillan Company, 1921. Pp. XVII + 289.

Sleight, W. G. *The Organization and Curricula of Schools.* VIII + 264.

D. New School Textbooks.

Ames, E. W., and Eldred, A. *Community Civics.* New York: MacMillan Company, 1921. Pp. XIV + 387.

Burch, H. R. *American Economic Life.* New York: MacMillan Company, 1921. Pp. XI + 533.

Hatfield, W. W. *Business English Projects.* New York: MacMillan Company, 1921. Pp. XV + 303.

Washburne, C. W. *Common Science.* New York: World Book Co., 1920. Pp. XV + 390.

The Multum in Parvo Atlas of the World. Chicago: A. J. Nystrom & Co., 1921. Pp. 46.

E. PUBLICATIONS OF UNITED STATES BUREAU OF EDUCATION.

Organization of State Departments of Education. Bureau of Education. Bulletin No. 46, 1920. Pp. 48.

Education for Highway Engineering and Highway Transport. Bureau of Education. Bulletin No. 42, 1920. Pp. 134.

Statistics of State Universities and State Colleges. Bureau of Education. Bulletin No. 87, 1919. Pp. 28.

Statistics of City School Systems, 1917-18. Bureau of Education. Bulletin No. 24. Pp. 477.

F. MISCELLANEOUS PUBLICATIONS.

COLLEGE TEACHERS OF EDUCATION. *Studies in Education.* Educational Monographs No. X. Maryland: King Bros., 1921. Pp. 79.

POWERS, S. R. *A History of the Teaching of Chemistry in the Secondary Schools of the United States Previous to* 1850. Minneapolis: University of Minnesota, 1920. Pp. 68.

The Preliminary Report of the Association Commission on the Organization of the College Curriculum. Chicago: Association of American Colleges Bulletin, 1921. Pp. 60.

ARNETT, T. *Teachers' Salaries in Certain Endowed Colleges and Universities in the United States.* New York: General Education Board, 1921. Pp. 42.

THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING. *Fifteenth Annual Report of the President and of the Treasurer.* New York, 1920. Pp. 171.